

RICE UNIVERSITY

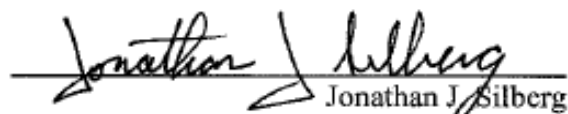
**Characterizing the tolerance of near infrared fluorescent bacterial phytochromes to
random backbone fission and circular permutation**


by
Naresh Pandey

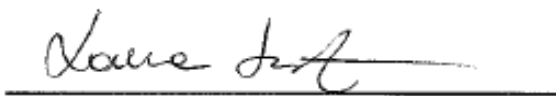
A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

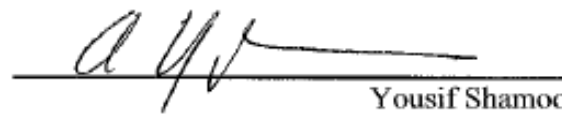
Doctor of Philosophy

APPROVED, DISSERTATION COMMITTEE


Jonathan J. Silberg
Associate Professor of BioSciences
Dissertation Chair


Edward P. Nikonowicz
Associate Professor of BioSciences


Laura Segatori
Associate Professor of Chemical and Biomolecular Engineering


Yousif Shamoo
Professor of BioSciences
Vice Provost for Research

HOUSTON, TEXAS
MAY 2016

Abstract

Characterizing the tolerance of near infrared fluorescent bacterial phytochromes to random backbone fission and circular permutation

by

Naresh Pandey

Protein fission, fusion, and circular permutation have been used to convert green fluorescent protein (GFP) family members into biosensors that dynamically report on cellular processes, ranging from protein expression and metabolite concentrations to protein solubility, protein-protein interactions, and ligand-binding. Unfortunately, GFP are unsuitable for deep tissue reporting in animal models because the wavelengths of light used with these reporters are highly absorbed by tissues. In contrast, near infrared fluorescent proteins (IFP and iRFP) derived from bacterial phytochrome proteins (BphP) are excited by light in the near-infrared spectrum (~700 nm, less absorptive) and are better suited for probing cellular processes within tissues.

To better understand the tolerance of BphP to various mutational lesions (fission, fusion, and circular permutation), I have subjected IFP to random backbone fragmentation and iRFP to circular permutation using transposase mutagenesis. Screening a library of split IFP for fluorescent variants yielded thirteen unique fragmented IFP and with parent like spectral properties. These two-fragment IFP all required assistance from associating proteins for maximal fluorescence. In all cases, the split IFP displayed AND gate logic behavior when the open reading frames (ORFs) encoding the different fragment were placed under distinct transcriptional regulation. In addition, screening a

library of circularly permuted iRFP led to the discovery of twenty seven variants with near infrared fluorescence. These variants arose from backbone fission in both PAS and GAF domains, although the brightest permuted iRFP variants initiated translation at residues near the domain linker and termini. Biochemical analysis revealed that permuted iRFP display similar oligomerization, quantum yield, and stability as native iRFP. These proteins also retained sufficient BV affinity to serve as reporters of gene expression in mammalian cells without the addition of exogenous BV.

These results demonstrate that knotted BphP retain the ability to fold as their contact order changes, suggesting that these proteins can be further developed as reporters of biological processes. The split IFP represent a suite of assays that will be useful for monitoring the dynamics of protein-protein interactions under conditions where split GFP do not yield strong signals. In addition, the circularly permuted iRFP should be useful for building molecular switches through domain insertion.

Acknowledgements

I would like to thank my advisor and mentor, Dr. Jonathan (Joff) Silberg, for giving me opportunity to work in his lab to pursue my Ph.D. I would also like to thank Silberg Lab, Matthews Lab members and Shirley Liu for helping me along the way. I thank my committee members (Dr. Shamoo, Dr. Nikonwicz and Dr. Segatori) for their insight and guidance through the years. Thanks to my collaborators Dr. Christopher Nobles, Dr. Anthony Maresso and Dr. Lynn Zechiedrich at Baylor College of Medicine for their help and letting me work in their labs. Special thanks to Barbara Nassif, Brianna Kuypers, Emily Thomas and Razan Alnahhas for helping to generate data for this thesis. I also thank my family for supporting me throughout this journey.

TABLE OF CONTENTS

List of tables.....	ix
List of illustrations	x
List of abbreviations	xiii
Chapter 1. Introduction	1
1.1 Genetically-encoded fluorescent proteins	1
1.2 The GFP protein family	1
1.2.1 Discovery of GFP	1
1.2.2 Applications of GFP as simple reporters	2
1.2.3 Discovery and engineering of additional colors	3
1.2.4 Introducing biosensing functions into GFP	4
1.2.5 Limitations of GFP-type reporters	6
1.3 The IFP family	6
1.3.1 Development of the first IFP	6
1.3.2 Discovery of iRFP	7
1.3.3 Advantages of IFP and iRFP	9
1.3.4 Bacterial phytochromes have a knotted topology	10
1.4 IFP/iRFP have potential as biosensors	13
1.5 Designing proteins using combinatorial experiments	13
1.5.1 Creating libraries of randomly fragmented proteins	15
1.5.2 Creating libraries of randomly circularly permuted proteins	17
1.6 Summary of my thesis research.....	19

Chapter 2. Materials and Methods	21
2.1 Library construction	21
2.1.1 Library of split IFP variants	21
2.1.2 Library of circularly permuted iRFP variants	24
2.2 High-throughput library screening	26
2.3 Whole cell fluorescence measurements	27
2.4 Vector construction	28
2.5 BV-dependence assay	29
2.6 Western immunoblot	29
2.7 Protein production and purification	30
2.8 Gel filtration chromatography	31
2.9 Native gel electrophoresis	31
2.10 Quantum yield determination	32
2.11 Extinction coefficient determination	32
2.12 Equilibrium unfolding	32
2.13 Mammalian tissue culture and flow cytometry	33
Chapter 3. IFP tolerance to random fission and rational fusion.....	34
3.1 Abstract	34
3.2 Introduction	35
3.3 Results	39
3.3.1 Combining IFP fission with fusion to peptides	39
3.3.2 IFP tolerates fission when fused to peptides	40
3.3.3 Split IFP fluoresce better at 23°C	40

3.3.4 Split IFP have similar BV dependence	42
3.3.5 Split IFP function as AND gates	46
3.3.6 Fragmented IFP require assistance for fluorescence	48
3.3.7 Combining CheA/CheY fusion with IFP fission	58
3.4 Discussion	63
Chapter 4. iRFP tolerance to random circular permutation	68
4.1 Abstract	68
4.2 Introduction	69
4.3 Results	72
4.3.1 Generation of permuted iRFP	72
4.3.2 Screening the library identified multiple variants	72
4.3.3 Permuted variants display low contact order	74
4.3.4 A majority of permuted iRFP fluoresce better at 37°C	77
4.3.5 Permuted iRFP have similar BV dependence as iRFP	79
4.3.6 Relationship between translation initiation and fluorescence	79
4.3.7 <i>In vitro</i> analysis of circularly permuted iRFP	85
4.3.8 Fluorescence in mammalian cells	94
4.4 Discussion	96
Chapter 5. Conclusions and future directions.....	101
5.1 Knotted IFP are sensitive to backbone fission	101
5.2 Split IFP should be useful for studying protein-protein interactions <i>in situ</i>	101
5.3 A generalizable method for discovering split protein that depend on protein-protein interactions	103

5.4 Split proteins will be useful for studying DNA minivector delivery <i>in situ</i>	104
5.5 Circularly permuted iRFP will be useful for domain insertion	106
5.6 Conclusions and Outlook	108
References	110

List Of Tables

Table 3.1 Spectral properties of fragmented IFP	44
Table 4.1 Spectral properties of circularly permuted iRFP	75
Table 4.2 <i>In vitro</i> properties of circularly permuted iRFP	89

List Of Illustrations

Figure 1.1 Reporting capabilities of GFP-type reporters.....	5
Figure 1.2 Structural comparison of GFP and IFP	8
Figure 1.3 Amino acid sequences alignment of IFP and iRFP	11
Figure 1.4 IFP contains a knot	12
Figure 1.5 IFP has limited reporting capabilities.....	14
Figure 1.6 Schematic illustration of IFP and iRFP fission and permutation	18
Figure 2.1 Scheme for creating libraries of fragmented IFP	22
Figure 2.2 Synthetic DNA used for the split IFP library construction	23
Figure 2.3 Scheme for creating a library of circularly permuted iRFP	25
Figure 3.1 Fragmented IFP created by random fission and rational fusion.....	38
Figure 3.2 Fluorescent fragmented IFP from the EK library.....	41
Figure 3.3 Fluorescence of fragmented IFP at 23°C	43
Figure 3.4 Effect of BV concentration on the fluorescence of fragmented IFP fused to IAAL-E3 and IAAL-K3 peptides	45
Figure 3.5 Two-input transcriptional regulation of fragmented IFP.....	47
Figure 3.6 140-EK displayed similar fluorescence to full-length IFP from either inducible promoter.....	49
Figure 3.7 IAAL-E3 and IAAL-K3 peptide removal diminishes IFP fragment complementation.....	50
Figure 3.8 Effect of biliverdin concentration on the fluorescence of fragmented IFP lacking fusion to any peptides or proteins	52

Figure 3.9 Effect of changing protein fusion on the levels of IFP fragments within cells	53
Figure 3.10 Calculated translation initiation rates for C-terminal fragments containing or lacking fusion to IAAL-K3	54
Figure 3.11 Removal of IAAL-E3 decreases IFP fragment complementation.....	56
Figure 3.12 CheA and CheY rescue IFP fragment complementation	57
Figure 3.13 Fluorescent fragmented IFP discovered in the AK library.....	60
Figure 3.14 Fragmented IFP discovered in the AY library	61
Figure 3.15 IFP displayed increased fluorescence at 23°C	65
Figure 4.1 Quality assessment of the circularly permuted iRFP library.....	73
Figure 4.2 Mapping of the iRFP tolerance to permutation and contact order	76
Figure 4.3 Comparison of circularly permuted iRFP variants fluorescence at 23°C and 37°C	78
Figure 4.4 Effect of BV concentration on <i>E. coli</i> fluorescence expressing circularly permuted iRFP	80
Figure 4.5 Trend of BV dependence among circularly permuted iRFP	81
Figure 4.6 Thermodynamic analysis of permuted iRFP translation initiation.....	83
Figure 4.7 Addition of an N-terminal affinity tag enhances whole cell fluorescence	84
Figure 4.8 Immunoblot analysis of permuted protein expression.....	86
Figure 4.9 Protein-normalized fluorescence in <i>E. coli</i>	87
Figure 4.10 Absorbance of circularly permuted iRFP	90
Figure 4.11 Effect of permutation on iRFP structure	91
Figure 4.12 Equilibrium unfolding of circularly permuted iRFP	93

Figure 4.13 Flow cytometry analysis of tissue culture cells without exogenous BV	95
Figure 5.1 Split IFP for gene delivery using DNA minivectors	105
Figure 5.2 Domain insertion using permuted iRFP	107

List Of Abbreviations

APC-Cy7, Allophycocyanin conjugated with cyanine dye

ARA, Arabinose

AY, CheA and CheY proteins

BphP, Bacteriophytochrome

BV, Biliverdin

CO, Contact order

DHFR, Dihydrofolate reductase

E. coli, *Escherichia coli*

EK, IAAL-E3 and IAAL-K3 peptides

FRET, Fluorescence resonance energy transfer

GAF, cGMP phosphodiesterase/adenyl cyclase/FhlA

GFP, Green fluorescent protein

GndHCl, Guanidium hydrochloride

GST, Glutathione S-transferase

HA, Hemagglutinin

IFP, Near-infrared fluorescent protein (from *Deinococcus radiodurans*)

IPTG, Isopropyl β -D-1-thiogalactopyranoside

iRFP, Near-infrared fluorescent protein (from *Rhodopseudomonas palustris*)

Kan^R, Kanamycin nucleotidyl transferase

kb, Kilo base

kDa, Kilo dalton

LB, Luria broth

MuA, Bacteriophage Mu transposase

ORF, Open reading frame

Ori, Origin of replication

PAS, Per/Arndt/Sim

PBS, Phosphate-buffered saline

PERMUTE, PERmutation Using Transposase Engineering

R1R2, Transposase recognition sequence

R2R1, Transposase recognition sequence

RBS, Ribosomal binding site

SDS-PAGE, Sodium dodecyl sulfate-polyacrylamide gel electrophoresis

λ_{em} , Emission wavelength

λ_{ex} , Excitation wavelength

Chapter 1

Introduction

1.1 Genetically-encoded fluorescent proteins

Fluorescent proteins, which emit light after absorbing electromagnetic radiation of lower wavelength, are an important part of life sciences because their production can be coupled to variety of biological processes like protein expression, protein solubility, protein dynamics and localization (Cabantous et al., 2005; Lippincott-Schwartz et al., 2001; Felmeier et al., 2000). Their discovery resulted in molecular tools that enabled scientists to continuously monitor biological processes like cellular pH, metal homeostasis, metabolite concentrations, protein solubility, and protein-protein interactions (Kneen et al., 1998; Chapleau et al., 2008; Sakaguchi et al., 2009; Cabantous et al., 2005; Hoff et al., 2009). These proteins were transformative for biology because they emit light when excited by specific wavelength providing a powerful imaging tool for spatial and temporal information on biological processes (Sun et al., 1999). Fluorescent proteins either form their chromophore autocatalytically or bind to a ligand that exhibits fluorescence (Hastings et al., 1996). These chromophores must reside inside the protective protein scaffold to fluoresce.

1.2 The GFP protein family

1.2.1 Discovery of GFP

In 2008, the Nobel Prize committee awarded Prize in Chemistry "for the discovery and development of the green fluorescent protein, GFP". The first fluorescent protein to

be discovered was *Aequorea victoria* GFP, which exhibits green fluorescence upon excitation with blue light (Shimomura et al., 1962). GFP is composed of 238 residues and forms its own chromophore through an oxidative reaction (Heim et al., 1994; Barondeau et al., 2003). The chromophore is generated by cyclization of three residues (Ser, Tyr, and Gly at positions 65-67), which form a p-hydroxybenzylidene-imidazolidinone (Heim et al., 1994). Shimomura's group was the first to identify the presence of GFP in the extracts of jellyfish (*Aequorea*) in 1962 (Shimomura et al., 1962). In 1992, Prasher and group successfully cloned and sequence the GFP gene (Prasher et al., Gene 1992). In 1994, Chalfie's group demonstrated that the GFP can be used as a marker of gene expression in *E. coli* and *C. elegans* (Chalfie et al., Science 1994), and Roger Tsien's group demonstrated that oxygen is required for the chromophore maturation and engineered another variant of GFP, with a blue colored variant paving the way for multiple genes expression reporters (Heim et al., 1994).

1.2.2 Applications of GFP as simple reporters

GFP and its derivatives with different spectral properties (called GFP herein) have been widely used as non-invasive reporters of different cellular behaviors (for review, see Chaudakov et al., 2010). They represent a powerful analytical tool that can be used to measure the expression of any gene in any organism where the GFP fluorescent signal can be monitored. Expression detection was initially accomplished by tagging the *mec-7* gene (codes for beta-tubulin) in *C. elegans* neurons (Chalfie et al., 1994). GFP has also been used to provide information on cellular localization and transport mechanism of different proteins in cells. For example, maltose binding protein was tagged with GFP to

study the export mechanism in periplasmic space of bacteria (Felmeier et al., 2000). In addition, GFP has been used to identify cells expressing specific proteins at a single cell level and to establish the variability in expression across a cell population (Longo et al., 2006). Furthermore, to determine the localization of proteins, GFP has been fused to either the N- and C-termini of proteins, and the cellular localization has been monitored using microscopy (Palmer et al., 2004). Fluorescent proteins are also useful for labeling specific organs and subcellular regions (Palmer et al., 2004). GFP can also be used as a genetic screen to quantify the activation different promoters across varying cell types (Ducrest et al., 2002).

1.2.3 Discovery and engineering of additional colors

The original GFP developed by Tsien and coworkers only allowed the study of a single biological reaction at a time. To overcome this limitation, a palette of GFP orthologs with different colors were developed that allow for the detection of multiple cellular events simultaneously (Heim et al., 1996). This was accomplished using directed evolution, where mutation and screening were used to identify mutations that alter the spectral properties of GFP without disrupting folding (Heim et al., 1996; Nagai et al., 2002). Since that time, GFP has been extensively engineered to create dozens of variants with different colors (and spectral characteristics), including enhanced green fluorescent protein ($\lambda_{\text{ex}} = 489 \text{ nm}$, $\lambda_{\text{em}} = 509 \text{ nm}$), blue fluorescent protein ($\lambda_{\text{ex}} = 402 \text{ nm}$, $\lambda_{\text{em}} = 457 \text{ nm}$), cyan fluorescent protein ($\lambda_{\text{ex}} = 434 \text{ nm}$, $\lambda_{\text{em}} = 477 \text{ nm}$), and yellow fluorescent protein or Venus ($\lambda_{\text{ex}} = 513 \text{ nm}$, $\lambda_{\text{em}} = 527 \text{ nm}$). In addition, gene sequencing in fluorescent organisms has revealed a palette of GFP orthologs with different colors,

including red variants from Corals (Matz et al., 1999; Fradkov et al., 2002; Bevis et al., 2002). To date, the farthest red-shifted GFP is mKate2 ($\lambda_{\text{ex}} = 588 \text{ nm}$, $\lambda_{\text{em}} = 633 \text{ nm}$).

1.2.4 Introducing biosensing functions into GFP

Protein reporters like GFP have transformed biological studies, although they are limited in their applications as biosensors. They require some modifications to function as reporters of pH, metal, and protein-protein interactions. Protein engineering has been used to create GFP that report on these additional reactions (Figure 1.1). For example, calcium signals can be dynamically monitored in cells through microscopy using changes in a FRET (Fluorescence resonance energy transfer) signal arising from BFP and GFP, which are covalently linked to calmodulin and the calmodulin binding peptide (Miyawaki et al., 1997). The FRET signal changes whenever free Ca^{2+} binds with the fusion protein, resulting in the localization and monitoring Ca^{2+} signal in the cell. A pH-sensitive GFP was developed by creating point mutations (F64L, S65T, Y66H) near the chromophore (Kneen et al., 1998; Llpois et al., 1998). Fluorescence and absorbance changes of the mutant GFP can be used as a pH indicator when expressed in the cell organelles, which can have different pH than cytosol. Metabolites can be sensed using GFP like inositol, (Sakaguchi et al., 2009) and metal clusters (Hoff et al., 2009). The former required insertion of a PH domain into the circularly permuted GFP while the latter fused fragments of GFP to the glutaredoxins. The change in the fluorescence intensity after expression of these proteins is attributed to metabolite binding. These are powerful tools for studying dynamic biological reactions in cells, but they are limited to conditions where GFP spectral probes can be used, which represents only a subset of cells in nature.

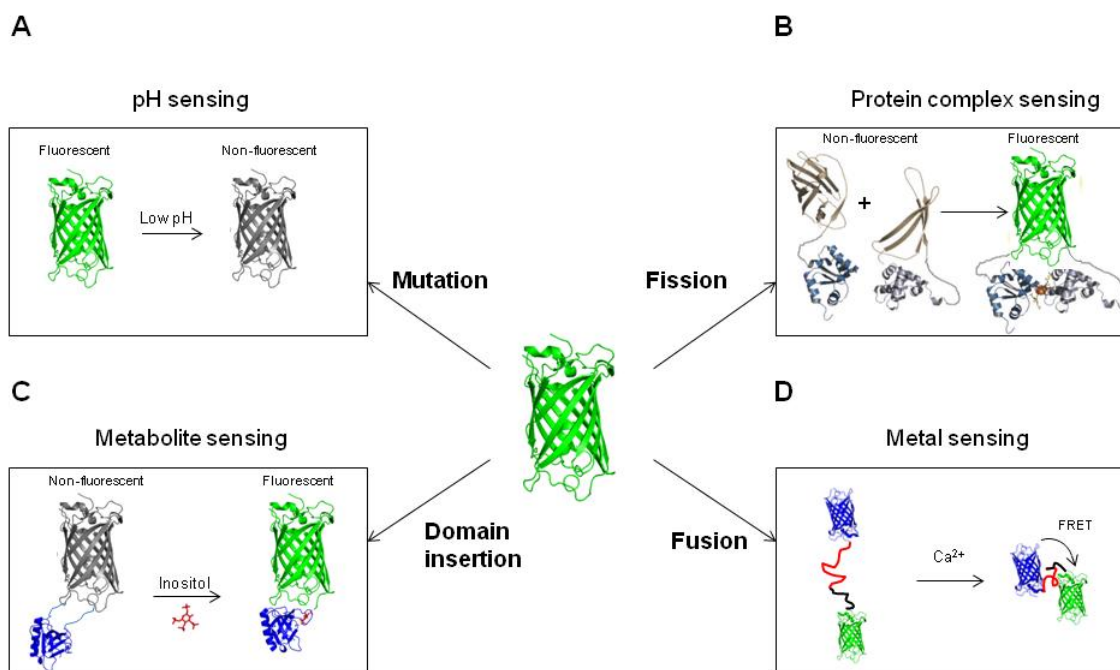


Figure 1.1 Reporting capabilities of GFP-type reporters. Examples of reporters include: (A) a pH-sensitive mutant which can report on intracellular pH conditions, (B) a split GFP that reports on protein-protein interactions when fused to interacting proteins like glutaredoxins, (C) a metabolite-sensing biosensor made up of a PH domain fused to GFP which reports on inositol concentrations, and (D) a metal sensing-GFP which is composed of a calcium-binding calmodulin, calmodulin binding peptide, BFP, and GFP fusion.

1.2.5 Limitations of GFP-type reporters

GFP-type reporters absorb and emit light within the visible spectrum range, which allows their use in cells that do not significantly absorb light in this range. However, this characteristic hinders our ability to use GFP for deep tissue imaging in animals (*e.g.*, gut or liver) and other complex conditions where the extinction of visible light is high (*e.g.*, soils and sediments). This occurs because these more complex settings display a strong absorption of visible light and exhibit autofluorescence (Jobsis et al., 1977; Shu et al., 2009). To overcome this challenge, a new class of fluorescent proteins called infrared fluorescent proteins (IFP or iRFP) was developed that absorb and emit light in near infrared region (Shu et al., 2009; Filonov et al., 2011). These proteins are better suited for animal tissues studies because tissues display lower autofluorescence at near infrared wavelengths compared with visible wavelengths and absorb near infrared wavelengths of light to a lesser extent than visible light (Weissleder et al., 2003).

1.3 The IFP family

1.3.1 Development of the first IFP

Phytochromes are red/far-red light photoreceptors that direct photosensory responses across the bacterial, fungal and plant kingdoms (for review, see Piatkevich et al., 2013). Early studies revealed that bacterial phytochromes display weak absorbance and fluorescence in the near infrared region (Lamparter et al., 2002; Giraud et al., 2005; Wagner et al., 2008). In 2005, a group from Wisconsin solved the structure of the chromophore-binding region of a bacterial phytochrome BphP from *Dienococcus radiodurans* (Wagner et al., 2005). This structure included the portion of the

phytochrome that is responsible for intrinsic fluorescence (Figure 1.2). In 2009, Roger Tsien and coworkers randomly mutated the chromophore binding domain of BphP and screened for variants with improved spectral properties (Shu et al., 2009). After multiple rounds of mutation and screening, they identified a variant that fluoresces four times brighter in the near infrared region, is stable over a wide pH range, encounters less photofatigue, and has an improved quantum yield and extinction coefficient. This mutant, named infrared fluorescent protein (IFP), contained 13 mutations.

1.3.2 Discovery of iRFP

The development of IFP yielded a genetically-encoded reporter with improved spectral properties compared to GFP for deep tissue imaging (Shu et al., 2009). This protein has been widely used for a variety of applications, including detection of malarial infection (Sigala et al., 2012), neural and tumor brain imaging (Yu et al., 2014), and tracking of a bacterial infection (Nobels et al., 2015). However, the original IFP developed by Tsien's group had several drawbacks. First, the photobleaching time for IFP is two times faster than the visible fluorescent proteins. Second, IFP requires the addition of exogenous biliverdin when used for mammalian tissue culture imaging. Third, IFP requires a high number of copies per cells to yield a signal that is sufficient for imaging due to its low overall molecular brightness. To overcome these limitations, *Rhodospseudomonas palustris* bacteriophytochrome RpBphP2 was engineered by Filonov and coworkers to create a second near infrared fluorescent proteins called iRFP (Filonov et al., 2011). IFP and iRFP display 39% sequence identity. iRFP also displays biliverdin-dependent fluorescence with similar biliverdin and biliverdin-protein absorbance peaks

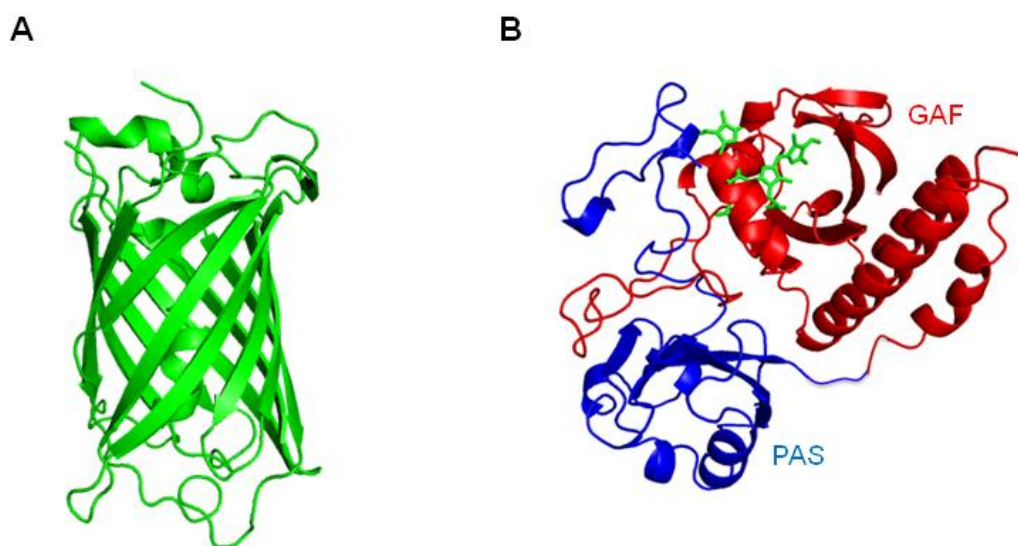


Figure 1.2 Structural comparison of GFP and IFP. **(A)** GFP is a barrel shaped protein comprised mostly of beta sheets. The GFP chromophore is formed through an autocatalytic process involving residues 65-67 in the middle of the barrel. **(B)** IFP/iRFP is a globular shaped protein with mixed alpha helices and beta sheets. They are comprised of two domains, including PAS and GAF domains. Biliverdin (green) is the chromophore, and it makes a thioether bond with a conserved cysteine residue at N-terminal. Images were created using PDB files 1GFL and 1ZTU.

near 391 nm and 700 nm, respectively. iRFP also displays a higher affinity for biliverdin, such that it is able to fluoresce in mammalian cells without addition of exogenous biliverdin. In addition, iRFP is a ten times brighter and more photostable than IFP. Since its development, iRFP has been used for a wide variety of applications, including deep tissue photoacoustic tomography (Filonov et al., 2012), localization of tumor margins in cancer models (Zhu et al., 2013), and as a marker for tumor growth in mice (Hock et al., 2014; Scherbakova et al., 2013).

1.3.3 Advantages of IFP and iRFP

Near-infrared light provides a window of opportunity for imaging in deep tissues because endogenous biological molecules do not absorb strongly in near infrared spectrum. Long wavelengths of light (650-950 nm) provide high signal to noise ratio in deep tissues imaging due to reduced tissue autofluorescence and absorption (Weissleder et al., 2003). Unfortunately, GFP type fluorescent reporters emit light within the visible spectrum, which hinders our ability to use them for deep tissue imaging in animals (*e.g.*, gut), due to the absorption of visible light (Jobsis et al., 1977). Fluorescent dyes were developed for to use in deep tissue but require extensive chemical preparation and lack efficient control of the fluorescence expression (Resch-Genger et al., 2008). To solve those problems, genetically encoded near-infrared fluorescent proteins (IFP and iRFP) were developed from bacterial phytochromes. These proteins allow for non-invasive monitoring of gene expression or disease progression in intact animals for longer periods of time. Despite their advantages in deep tissue imaging, both iRFP and IFP have low

quantum yields (6%) compared to GFP (60%). These proteins also dimerize when expressed in high concentrations, which might impede their usage as a reporter.

1.3.4 Bacterial phytochromes have a knotted topology

The structure of BphP, which was mutated to create IFP, was solved using crystallography in 2005 (Wagner et al., 2005). This structure revealed two distinct domains, including PAS and GAF domains. The biliverdin chromophore is covalently linked to the 24th residue (cysteine) of the IFP PAS domain through a thioether linkage (Figure 1.3). Within the structure, this chromophore is covalently linked to the PAS domain but buried within the GAF domain. BphP also contains a trefoil knot. This knot (Figure 1.4) arises because the first 33 residues must pass through a C-terminal loop (residues 225-257) to generate the native structure (Wagner, et al., 2005). Knotted proteins are thought to have lengthy folding times due to topological barrier of a knot during translation and folding (King et al., 2010). They have two major steps in folding. First, they must form a twisted native loop. Second, they must thread one of the termini through the loop (Sulkowska et al., 2012). In case of IFP, the loop region is located in the C-terminal domain so threading can only happen after the protein is synthesized by the ribosome, which could limit the rate at which this protein can mature. Knotted proteins have been proposed to have advantages. There is evidence that knots can provide extra mechanical stability to proteins (Sulkowska et al., 2008). *In vitro* experiment performed with a knotted protein showed spontaneous formation of the knotted topology without the any aid of molecular chaperones. However the rate of the knot formation was reduced in the absence of molecular chaperones (Mallam et al., 2011). For IFP, no previous


```

      10      20      30      40      50      60
IFP  MSRDPLPFFP PLYLGGPEIT TENQEREPIH IPGSIQPHGA LLTADGHSGE VLQMSLNAAT
iRFP -----M TEGSVARQPD LLTQDDEPIH IPGAIQPHGL LLALAAD-MT IVAGSDNLPE
      1      11      21      31      40      50

      70      80      88      98      108      118
IFP  FLGQEPTVLR GQTLAALLPE QWP--ALQAA LPPGCPDALQ YRATLDWPAA GHLSLTVHRV
iRFP  LTGLAIGALI GRSAADVFDSETHNRLTIAL AEPGAAVGAP ITVGFTMRKD AGFIGSWHRH
      60      70      80      90      100      110

      128      138      145      155      165      175
IFP  GELLILEFEP TEAWDSTGPH ALR---NAMS ALESAPNLRA LAEVATQTVR ELTGFDRVML
iRFP  DQLIFLELEP PQRDVAEPQA FFRRTNSAIR RLQAAETLES AAAAAQEVK KITGFDRVMI
      120      130      140      150      160      170

      185      195      205      215      225      235
IFP  YKFAPDATGE VIAEARREGL HAF LGHRFPA SHIPAQARAL YTRHLLRLTA DTRAAAVPLD
iRFP  YRFASDFSGE VIAEDRCAEV ESKLGLHYPA STVPAQARRL YTINPVRIIP DINYRPVPVT
      180      190      200      210      220      230

      245      255      265      275      285      295
IFP  PVLNPQTNAP TPLGGAVLRA TSPMHMQFLR NMGVGSLSV SVVVGGQLWG LIACHHQTPY
iRFP  PDLNPVTGRP IDLSFAILRS VSPVHLEFMR NIGMHGTMSI SILRGERLWG LIVCHHRTPY
      240      250      260      270      280      290

      305      315      325
IFP  VLPPDLRTTL EYLGRELSEQ VQVKEALEHH HHHH
iRFP  YVDLDGRQAC ELVAQVLAWQ IGVME-----
      300      310      316

```

Figure 1.3 Amino acid sequences alignment of IFP and iRFP. Residues in the PAS domain (blue), interdomain linker (gray) and GAF domain (red) are colored differently. Underlined cysteines bind biliverdin and make a covalent thioether linkage with biliverdin. Residues highlight in yellow contact biliverdin.

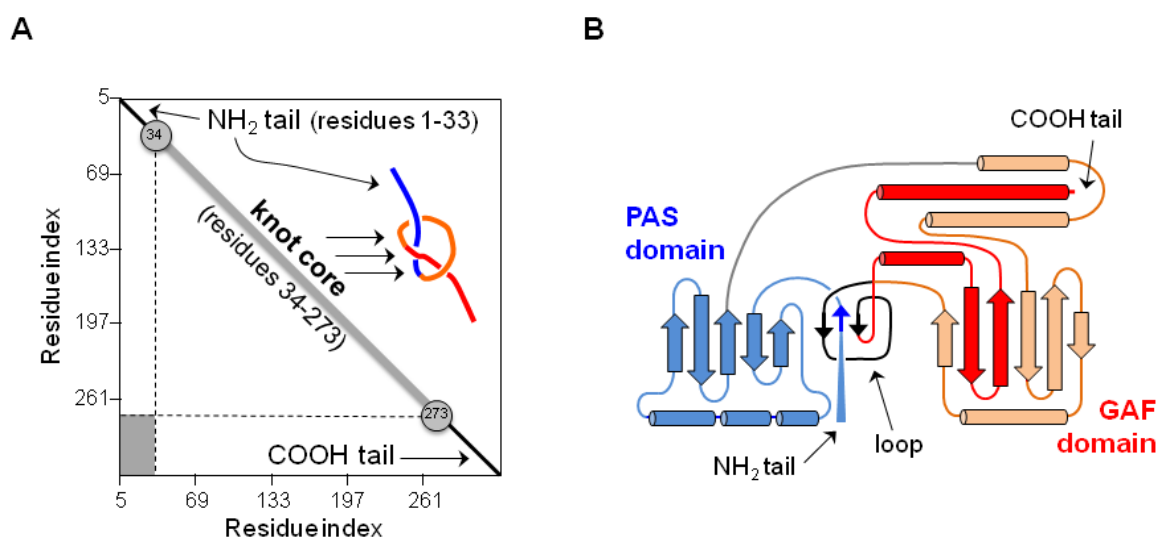


Figure 1.4 IFP contains a knot. (A) The plot shows knot features obtained from knotprot (<http://knotprot.cent.uw.edu.pl/>). IFP contains a trefoil knot that forms when the two tails interlace. The N-terminal tail is comprised of residues 1-33, the knot core made up of residues 34-273 and the C-terminal tail is comprised of residues 274-321. (B) IFP topology diagram showing a knot formation when first 33 residues (NH₂ tail) of PAS domain passed through C-terminal residues 225 to 257 (loop) of GAF domain.

experiments have been performed to probe the importance of the knot. Also, it is unclear previous experiments have been performed to probe the importance of the knot. Also, it is unclear how knot will tolerate to backbone fission and circular permutation. This thesis provides the first insight into knotted protein tolerance to random fission and circular permutation.

1.4 IFP/iRFP have potential as biosensors

GFP has been extensively engineered through directed evolution to create biosensors for biotechnological applications. By combining different mutational strategies (point mutation, recombination, splitting, fusion, and domain insertion) with screens, GFP have been identified that are capable of pH sensing, protein complex sensing, metal sensing and metabolite sensing (see Figure 1.1). In contrast, similar biosensors have not yet been developed using fluorescent bacterial phytochromes. For this reason, there is a need to engineer IFP to make new classes of biosensors for systems and synthetic biology that are compatible with measurements in tissues where GFP biosensors yield poor signals. However, prior to this thesis research, it was unclear if the knotted protein topology in BphP would tolerate backbone fission and circular permutation to the same extent as GFP. Generating split IFP and circularly permuted iRFP will allow researchers to create biosensors or applications suited for deep tissue (Figure 1.5). Circularly permuted iRFP can be used for domain insertion for generating allosteric switches and also can be used as fusion protein for building sensors.

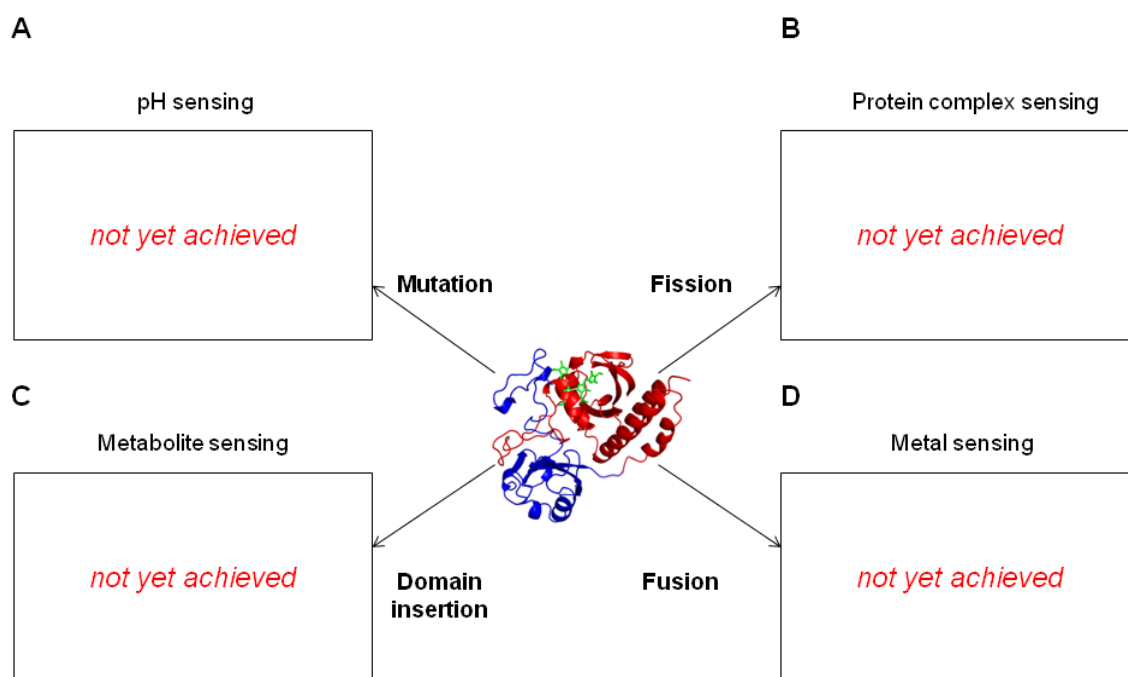


Figure 1.5 IFP has limited reporting capabilities. IFP has the potential to be engineered into biosensors with a range of functions, including: (A) pH sensing, (B) protein complex sensing, (C) metal sensing, and (D) metabolite sensing.

1.5 Designing proteins using combinatorial experiments

Laboratory evolution can be used to study where proteins tolerate different types of mutational lesions. In laboratory evolution, a library of protein variants is created and then screened (or selected) to examine the effects of sequence changes on protein structure and function (for review, see Yuan et al., 2005). Sequence diversity in libraries is created using a variety of approaches. These include random mutation, homologous recombination, backbone fission, and circular permutation. Studies with random backbone fission have shown that some proteins tolerate a large number of backbone fission sites, i.e., the fragments arising from fission cooperatively self assemble to function like a full-length protein (Segall-Shapiro et al., 2011). In addition, studies examining domain insertion have revealed that hybrid proteins can be generated by inserting one protein into the backbone of a second protein; these hybrid proteins can retain the function of the parent proteins and/or display novel switching activity (Giraldez et al., 2005). Studies examining protein tolerance to random circular permutation have shown that some proteins retain parent-like function even though their contact order has changed (Mehta et al., 2012). Laboratory evolution can also be used to create protein variants with properties distinct from those observed in nature, which are useful for biotechnology applications. Prior to the work described in this thesis, the extent to which IFP tolerated random backbone fission and iRFP tolerated circular permutation had not been reported.

1.5.1 Creating libraries of randomly fragmented proteins

Proteins are robust macromolecules, which can be quite tolerant to the mutational that

they experience in nature. Surprisingly, proteins that experience backbone fragmentation frequently reassemble *in vivo* and function (Shiba et al., 1992; Landro et al., 1993). Using laboratory evolution, various proteins have been subjected to fission for biosensor purposes. For example split GFP has been used to study protein-protein interactions (Barnard et al., 2008), where non-fluorescent GFP fragments were attached to the proteins of interest. An interaction between the proteins of interest resulted in the assembly of GFP fragments and a signal as GFP matured into a fluorescent protein. Split DHFR (Dihydrofolate reductase) has also been used as a protein-fragment complement assay (Remy et al., 2007). With DHFR, two proteins of interest were fused to DHFR fragments, and the physical interaction between proteins of interest resulted in a functional DHFR. Split adenylate kinase can also complement the bacterial growth (Nguyen et al., 2008). These adenylate kinase fragments complemented a growth defect of an *E. coli* strain whose adenylate kinase is non functional at high temperatures. T7 RNA polymerase can also function as a split protein AND gate (Shis et al., 2013). Cells expressing fragments of T7 RNA polymerase transcribe GFP when it is placed under control of the T7 promoter. The effects of stability on tolerance to fission has also been examined and revealed that thermophilic proteins tolerate random backbone fission to a greater extent than mesophilic proteins. In fact, the Silberg lab showed that more than 40% of all split variants of a thermophilic adenylate kinase retained parent-like function (Segall-Shapiro et al., 2011), while only 6% of the possible split variants of a mesophilic adenylate kinase retained detectable activity under similar assay conditions.

Although DNA synthesis has become cheap, it can still be arduous to use oligonucleotide-directed mutagenesis to make all of the open reading frames that encode

the different possible two-fragment versions of a natural protein. Transposon mutagenesis has appeared as a simple method for building libraries of randomly split proteins (Segall-Shapiro et al., 2010). In this method, a transposon is randomly inserted into an expression vector containing the gene of interest. The transposon is then replaced by a synthetic construct to incorporate the regulatory elements needed to transcribe and translate the two open reading frames created by gene fission. Using this approach, any protein can be subjected to backbone fission and screened or selected for the functional variants. In this thesis, I describe how IFP has been subjected to backbone fission using transposase mutagenesis. Prior to this thesis, it was not known that if IFP could tolerate backbone fission within the PAS and GAF domains or within the interdomain linker (Figure 1.6A), because no previous studies had examined IFP tolerance to random backbone fission.

1.5.2 Creating libraries of randomly circularly permuted proteins

Circular permutation has been extensively used in protein engineering to study protein folding (Haglund et al., 2008), improve the protein functions (Yu et al., 2010), alter substrate or ligand binding properties (Maatta et al., 2008), alter oligomerization (Yang et al., 1993), create allosteric regulation (Ostermeier et al., 2005) and develop proteins with novel functions by domain insertion (Baird et al., 1999). Despite the generation of a large number of circularly permuted protein variants, there is still lack of rules governing this design. Often a rational approach is used to create permuted proteins by targeting flexible regions or loops in protein structures (for proteins where structural information is available). Library methods have the advantage over rational design that they sample all of the possible sites to obtain the variants that are tolerant to circular

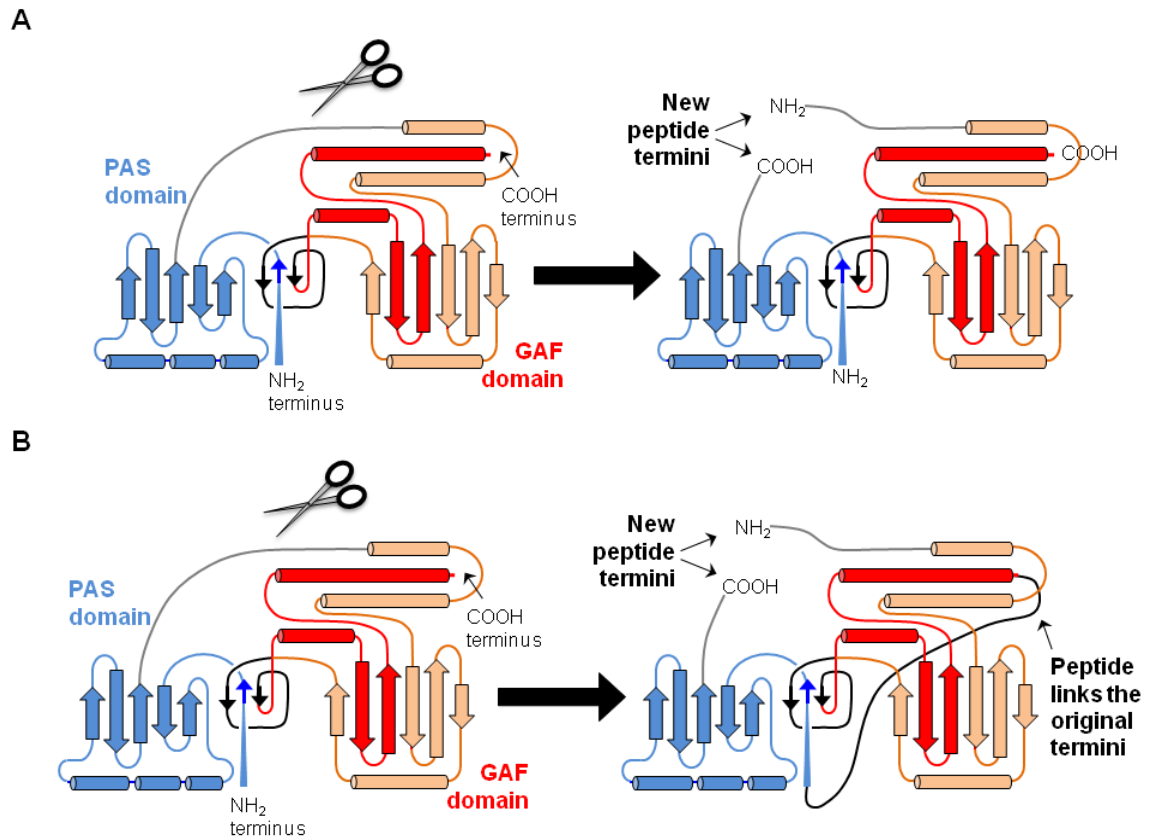


Figure 1.6 Schematic illustrations of IFP and iRFP fission and permutation. **(A)** Some sites in the protein backbone are more tolerant to fission than others. During backbone fission, IFP is randomly cleaved at different backbone locations to create two fragments. Each fragment has its own N- and C-termini, and when translated they associate and form a folded IFP. **(B)** Some sites in the protein backbone are predicted to be more tolerant to circular permutation than others. During circular permutation, the original termini are linked with flexible peptide linker and new peptide termini are created by fission elsewhere in the backbone.

permutation. Methods have been developed to create library of random permuted proteins (Graf et al., 1996; Hennecke et al., 1999; Guntas et al., 2004; Hida et al., 2010). However, the libraries generated using those methods contain deletions and duplications. The Silberg Lab developed a method to create circularly permuted proteins using transposon-based mutagenesis, which can rapidly generate libraries of permuted proteins without deletions and large duplications (Mehta et al., 2012). In this thesis, I describe an improvement to this method, and I describe the first efforts to use this new approach to study the effects of circular permutation on a knotted protein. The goal of this thesis was to see where iRFP tolerates backbone fission upon circular permutation (Figure 1.6B).

1.6 Summary of my thesis research

Near infrared fluorescent proteins can be useful for whole body imaging in mammals. However, the IFP toolbox only contains simple reporters of gene expression, which have not yet been diversified into a toolkit of biosensors similar to GFP (Tainaka et al., 2010; Lindenburg et al., 2014). I hypothesized that: (i) fragmented and circularly permuted IFP can be generated using directed evolution and (ii) these variants can be used for engineering novel IFP based biosensors. In Chapter 2, I describe the methods used to perform my thesis research. In Chapter 3, I describe my studies examining IFP tolerance to backbone fission and my efforts characterizing fragmented IFP dependence upon protein-protein interactions. For these studies, I generated a library of split IFP using transposase mutagenesis and screened for functional variants. These split variants were tested with different protein-protein interactions to determine if there is a correlation between binding affinity and fluorescence. In Chapter 4, I describe my studies examining where iRFP tolerates backbone fission arising from circular permutation. Library

screening identified over two dozen variants that display near infrared fluorescence. Purified permuted iRFP showed biochemical properties similar to wild type. In addition, permuted iRFP displayed near infrared fluorescence in mammalian tissue culture without the addition of exogenous biliverdin. In Chapter 5, I review my findings and describe how my experiments have expanded the IFP/iRFP toolbox. In addition, I discuss applications and future paths others can take that would build upon the discoveries presented herein.

Chapter 2

Materials and Methods

2.1 Library construction

2.1.1 Library of split IFP variants

A library of randomly fragmented *ifp* genes was constructed as illustrated in Figure 2.1 using pGEX2TK-IFP1.4, a pGEX-2TK-derived vector that expresses IFP1.4 (Shu et al., 2009) from a *tac* promoter with glutathione S-transferase (GST) fused to its N-terminus and a hemagglutinin (HA) tag fused to its C-terminus. A transposon containing NotI restriction sites proximal to both ends was inserted into pGEX2TK-IFP1.4 by incubating 310 ng pGEX2TK-IFP1.4, 100 ng M1-Kan^R transposon (Thermo Scientific), and 1 U HyperMu MuA transposase (Epicentre Biotechnologies) in a 20 µL reaction containing HyperMu buffer for 14 hours at 37°C. Reactions were terminated by adding HyperMu Stop Solution (1 µL) and incubating each reaction at 70°C for 10 min. Total DNA was purified, electroporated into *Escherichia coli*, spread onto multiple Luria Broth (LB)-agar plates containing 25 µg/mL kanamycin, and grown at 37°C overnight. Total plasmid DNA was purified from colonies obtained on plates to obtain the Mu-insertion library. The ensemble of purified vectors was digested using restriction enzymes (BamHI and EcoRI) that cut at sites flanking the IFP gene. Then, agarose electrophoresis was used to purify the *ifp*-transposon hybrids (2.1 kb) away from the other DNA fragments (1 kb *ifp* alone, 5 kb vector backbone, and 6.1 and 7.2 kb vector backbones containing one or two 1.1 kb transposons). The *ifp*-transposon hybrids were cloned back into pGEX2TK to

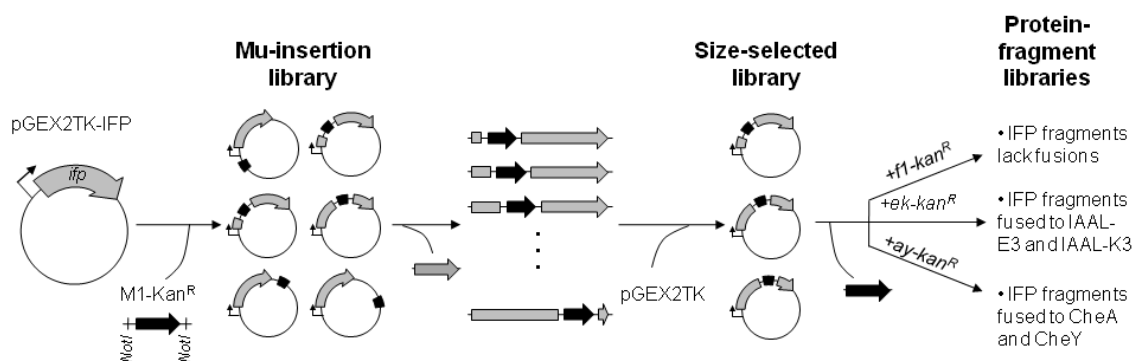


Figure 2.1 Scheme for creating libraries of fragmented IFP. Library construction involves four steps, including: (i) random insertion of transposons into the IFP genes using MuA, (ii) purifying IFP genes with inserted transposons away from IFP genes lacking inserts, (iii) subcloning the ensemble of IFP-transposon hybrid genes into an expression vector, and (iv) subcloning synthetic DNA encoding different pairs of interacting proteins in place of the transposons. The last step can be performed using synthetic DNA that encode for interacting peptides (IAAL-E3 and IAAL-K3), proteins (CheA and CheY), or lacking additional ORFs as shown in Figure 2.2.

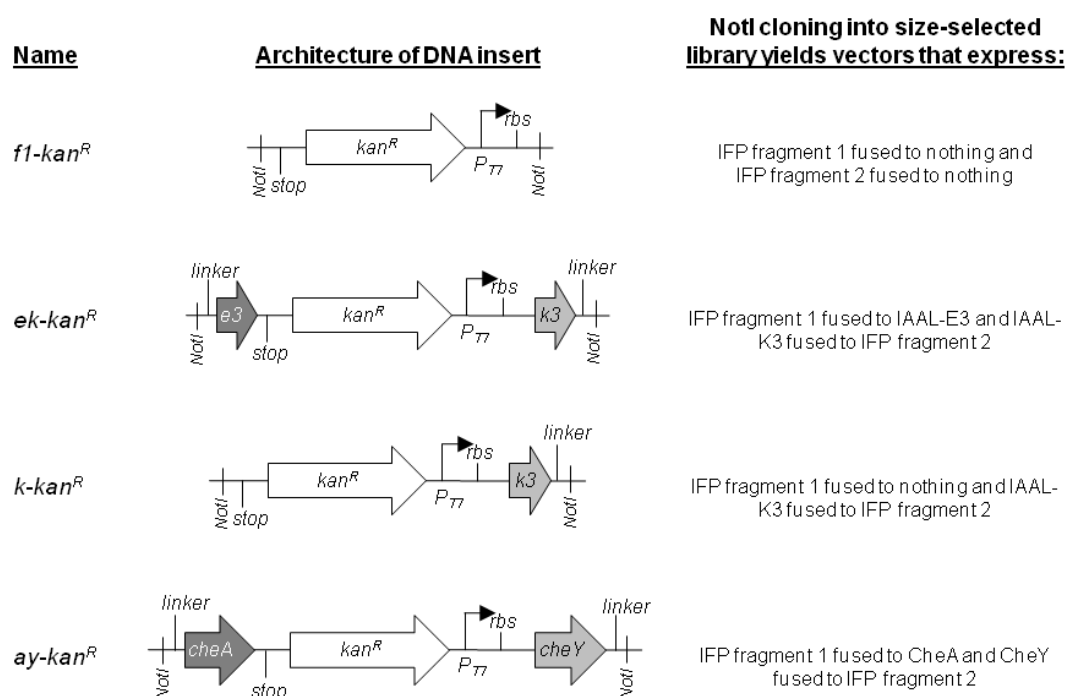


Figure 2.2 Synthetic DNA used for the split IFP library construction. The architecture of the four DNA used to create vectors that express fragmented IFP. Upon insertion into IFP genes, all of these DNA terminate expression of the IFP fragment that precedes the original transposon insertion site and initiate expression of the IFP fragment that follows the ORF. *f1-kan^R* does not amend associating peptides or proteins to either IFP fragment. *ek-kan^R* fuses IAAL-E3 to the end of the IFP fragments that precede the fragmentation site and IAAL-K3 to the beginning of the IFP fragments that follow the fragmentation sites. *k-kan^R* fuses IAAL-K3 to the beginning of the IFP fragments that follow the fragmentation sites and nothing to the fragment that precedes the fragmentation site. *ay-kan^R* fuses CheA to the end of the IFP fragments that precede the fragmentation site and CheY to the beginning of the IFP fragments that follow the fragmentation sites.

create a size-selected library (pGEX2TK-IFP-SS). DNA inserts shown in Figure 2.2 (*f1-kan^R*, *ek-kan^R*, and *ay-kan^R*) were subcloned in place of the transposon within pGEX-IFP-SS to create three different libraries of vectors that express randomly fragmented IFP, including: (i) the split IFP library that expresses fragmented IFP without fusion to other proteins, (ii) the EK library, which expresses the N- and C-terminal IFP fragments as fusions to the IAAL-E3 and IAAL-K3 peptides, and (iii) the AY library, which expresses the N- and C-terminal IFP fragments as fusions to *Thermotoga maritima* CheA (P2 domain; residues 175-264) and CheY. All steps involving bacterial transformations used MegaX DH10B competent cells (Life Technologies) and yielded lawns of colonies (>50,000 per transformation) whose numbers were always >10-fold larger than the number of variants in each library. Each library contained similar sequence diversity (1,924 variants), which is determined by the number of sites in the IFP gene where a NotI site is inserted (962) by the transposase MuA and the number of orientations (2) that each synthetic DNA (*f1-kan^R*, *ek-kan^R*, and *ay-kan^R*) can be integrated into the IFP gene.

2.1.2 Library of circularly permuted iRFP variants

Circularly permuted *iRFP* genes were constructed with modification of a previously described method (Mehta et al., 2012) as illustrated in Figure 2.3. NotI restriction sites flanking *iRFP* genes were amplified using primers from pBAD/His-B-iRFP (addgene). NotI flanked *iRFP* genes were digested with NotI, agarose gel purified and 60 ng of purified DNA was self ligated for 16 hours at 16°C to obtained a circularized *iRFP* genes. 100 ng of BglII linearized transposon (containing tac promoter, pBR322 origin of replication and kanamycin resistance cassette) was inserted into purified circularized

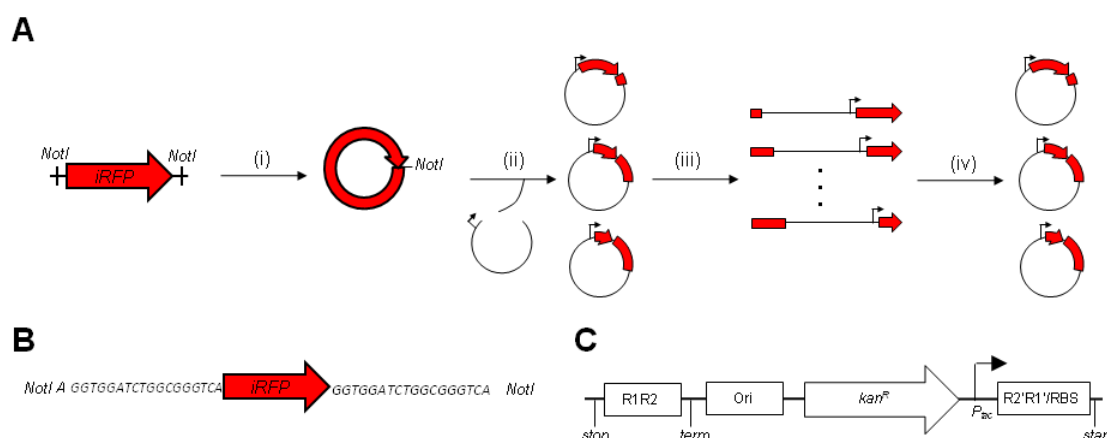


Figure 2.3 Scheme for creating a library of circularly permuted iRFP. **(A)** In this method, (i) NotI flanked *iRFP* genes are self ligated after NotI restriction digestion to create a circular *iRFP* genes containing twelve residues linker to join original N- and C-termini. (ii) A synthetic permuteposon P4 (black open circle) is randomly inserted into the circular *iRFP* genes by transposase (MuA) mutagenesis (iii) *iRFP* genes harboring an integrated permuteposon are excised from vectors containing mutiple permuteposon insertion using NotI digestion and (iv) these genes are self-ligated to create a library of vectors that express the different circularly permuted iRFP variants. **(B)** The target gene lacks a stop codon and is flanked by identical restriction sites (NotI) with six glycine rich linkers (GGSGGS), which become fused after self ligation and encode the linker that connects the original N- and C-termini of iRFP. An additional adenine was inserted between the initial NotI site and the *linker* gene to keep the linker in frame upon permutation. **(C)** The permuteposon P4 contains a stop codon (stop), MuA recognition site (R1R2), a terminator (term), a pBR322-derived origin of replication (ori), an IPTG inducible tac promoter (P_{tac}), a *kan^R* selectable marker, a hybrid MuA recognition site (R2'R1') with ribosomal binding site (RBS) and a start codon (start).

iRFP genes by incubating with 1 U HyperMu MuA transposase in a 20 μ L reaction containing HyperMu buffer for 16 hours at 37°C and terminated by incubating each reaction at 70°C for 10 min. Purified DNA electroporated into *E. coli* and spread onto LB-agar plates containing 25 μ g/mL kanamycin, and grown at 37°C overnight. Plasmid DNA was purified from colonies obtained on plates, linearized using NotI and run on agarose electrophoresis to purify the iRFP-transposon hybrids (2.9 kb). The iRFP-transposon hybrids were self ligated for 16 hours at 16°C to create a closed vector for expression. DNA was purified, electroporated into *E. coli*, spread onto LB-agar plates containing 25 μ g/mL kanamycin, and grown at 37°C overnight. All bacterial transformations involved use of MegaX DH10B competent cells and yielded lawns of colonies (>50,000 per transformation) which were >20-fold larger than the number of variants in each library.

2.2 High-throughput library screening

E. coli Rosetta 2 DE3 (Novagen) or BL21 Star DE3 (Life Technologies) or XL1 Blue transformed with each library were spread onto LB-agar plates containing 100 μ g/ml ampicillin or 50 μ g/mL kanamycin and incubated overnight at 37°C. Single colonies from these plates were arrayed into 96-well deep well plates containing 200 μ L LB and 100 μ g/mL ampicillin (or 25 μ g/mL kanamycin) and grown for 18 hours at 37°C while shaking at 250 rpm. Stationary phase cultures were diluted 4x by adding LB containing 100 μ g/mL ampicillin or (25 μ g/mL kanamycin). After growing for 1 hour at 37°C, isopropyl β -D-1-thiogalactopyranoside (IPTG) and BV (Frontier Scientific) were added to final concentrations of 0.5 mM and 80 μ M, respectively. Deep well plates were

incubated at 23°C while shaking at 250 rpm for 18 hours in the dark. A fraction (150 μ L) of each culture was transferred to clear polystyrene 96-well flat bottom plates (Corning) or 96-well v bottom plates (Rainin), cells were pelleted by centrifuging plates at 3,000 \times g for 5 minutes, supernatant containing excess BV was removed, and whole cell fluorescence ($\lambda_{\text{ex}} = 684$ nm; $\lambda_{\text{em}} = 695$ -720 nm or $\lambda_{\text{ex}} = 690$ nm; $\lambda_{\text{em}} = 700$ -725 nm) was measured using a Tecan M1000 plate reader. IFP variants with fluorescence (703 to 714 nm) that was $>3\sigma$ higher than the signal obtained from cells lacking the IFP gene were sequenced and given names that correspond to the IFP residue at the end of the first fragment followed by letters that describe the proteins fused to the termini created by fragmentation, where EK designates IAAL-E3 and IAAL-K3 and AY designates CheA and CheY. For iRFP, emission data was normalized to absorbance in each well and variants with fluorescence (705 to 720) that was $>5\sigma$ higher than the signal obtained from cells lacking the iRFP gene were sequenced and given names that correspond to the iRFP residue appeared after start codon.

2.3 Whole cell fluorescence measurements

Vectors encoding fragmented IFP or circularly permuted iRFP were transformed into *E. coli* BL21 Star DE3 or XL1 Blue, and individual colonies were used to inoculate LB cultures containing 100 μ g/mL ampicillin or 50 μ g/mL kanamycin. After 16 hour at 37°C and 250 rpm, cells (1 mL) were harvested by centrifugation and used to inoculate a fresh 5 mL LB culture containing 0.5 mM IPTG, 80 μ M BV, and 100 μ g/mL ampicillin or 50 μ g/mL kanamycin. Cells were grown for 5 hours at the indicated temperatures and 250 rpm in the dark, washed with 25% glycerol (1 mL), and resuspended in 25% glycerol (1

mL). Whole cell absorbance (600 nm) and fluorescence ($\lambda_{\text{ex}} = 684 \text{ nm}$, $\lambda_{\text{em}} = 695\text{-}800 \text{ nm}$ or $\lambda_{\text{ex}} = 690 \text{ nm}$, $\lambda_{\text{em}} = 700\text{-}800 \text{ nm}$) were acquired from samples arrayed in flat bottom 96-well plates using a Tecan M1000 plate reader. Emission data was normalized to absorbance in each well, and data reported represent the average of three or more colonies for each sample with four replicates measured of each within 96 well plates. The vector that expresses full-length IFP (pGEX2TK-IFP1.4) and the same vector lacking the IFP gene were used as frames of references for fragmented IFP signals. For iRFP, vector that expresses full-length pBAD/His-B-iRFP and the transposon vector lacking the iRFP gene were used as frames of references for permuted iRFP signals.

2.4 Vector construction

To create pairs of vectors for analyzing regulated expression of each fragmented IFP, N-terminal IFP fragments were cloned into pQE80 $\Delta\text{cm}^{\text{R}}$, a plasmid with an IPTG-inducible T5 promoter, and C-terminal IFP fragments were cloned in place of T7 RNA polymerase in the plasmid pTara (Wycuff et al., 2000), which has an arabinose-inducible P_{BAD} promoter. pQE80 $\Delta\text{cm}^{\text{R}}$ was created by deleting the chloramphenicol resistance gene from pQE80 (Qiagen). All other expression vectors were generated by subcloning different NotI-flanked DNA inserts in place of the NotI-flanked inserts used to build vectors in each library. Expression vector for circularly permuted variants were created by cloning into pBAD/His-B vector under arabad promoter with His-tag fused to N-terminus of the circularly permuted variants. For mammalian expressions circularly permuted variants were created by cloning into pEGFP-C1 vector under cytomegalovirus promoter with HA tag fused to C-terminus of the variants.

2.5 BV-dependence assay

The BV dependence of whole cell fluorescence for IFP split variants were measured using protocol of section 2.3, except the concentration of BV was varied. For iRFP variants, vectors encoding circularly permuted iRFP were transformed into *E. coli* (XL1 Blue), and individual colonies were used to inoculate LB cultures containing 100 µg/mL ampicillin. After 16 hour at 37°C and 250 rpm, cells (2 mL) were harvested by centrifugation and used to inoculate a fresh 8 mL LB culture containing 1 mM Arabinose, and 100 µg/mL ampicillin. 1 mL culture added into each well in deep well plates. Various amounts of BV were added to individual wells and incubated at 37°C while shaking at 250 rpm for 5 hours in the dark. For each well, a fraction (200 µL) of each culture was transferred to clear polystyrene 96-well V bottom plates (Rainin), cells were pelleted by centrifuging plates at 3,000 x g for 5 minutes, supernatant containing excess BV was removed, and whole cell absorbance (600 nm) and fluorescence ($\lambda_{\text{ex}} = 690 \text{ nm}$; $\lambda_{\text{em}} = 700\text{-}800 \text{ nm}$) was measured using a Tecan M1000 plate reader.

2.6 Western immunoblot

E. coli expressing the different fragmented IFP (or circularly permuted iRFP) were grown as described for fluorescent analysis, harvested by centrifugation, and resuspended to identical optical densities. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was carried out under reducing conditions using NuPAGE 12% Bis-Tris Gels (Life Technologies) and MOPS SDS running buffer and transferred to Protran nitrocellulose membrane (Whatman) using a TE 22 Mini Tank Transfer Unit (GE Healthcare). After washing the nitrocellulose paper in TBST buffer (100 mM Tris pH 7.5,

150 mM NaCl, 0.1% Tween 20) for 5 minutes and blocking for 1 hour with 10% dry milk in TBST, the membranes were incubated for 1 hour with either GST rabbit (Millipore) or hemagglutinin Ab-1 (NeoMarkers) polyclonal antibodies at dilutions of 1:10,000 in TBST. The nitrocellulose was then incubated for 1 hour in TBST with secondary antibody, goat anti-rabbit IgG conjugated to peroxidase conjugate (Calbiochem), at a dilution of 1:10,000. Signals were detected using the ECL western blotting substrate (GE Healthcare) according to the manufacturer's protocol. For iRFP His-tag polyclonal antibody (Qiagen) at dilution of 1:1000 in TBST was used as primary antibody and goat anti-rabbit IgG conjugated to peroxidase conjugate (Calbiochem) at dilution of 1:1000 was used as secondary antibody.

2.7 Protein production and purification

E. coli (JW2509-2) were cotransformed with hemeoxygenase plasmid (pSR34-Bvd) and iRFP variants expressed from pBAD, spread onto LB-agar plates containing 100 µg/mL streptomycin and 100 µg/mL ampicillin and incubated overnight at 37°C. Single colonies were used to inoculate fresh 50 mL LB containing 100 µg/mL streptomycin and 100 µg/mL ampicillin for overnight growth. A 1 mL overnight culture was used for inoculating fresh 50 mL LB containing 100 µg/mL streptomycin and 100 µg/mL ampicillin. After 3.5 hrs at 37°C and 250 rpm, cells (10 mL) were harvested by centrifugation and used to inoculate a fresh 1 L LB culture containing 100 µg/mL streptomycin and 100 µg/mL ampicillin. The culture was grown at 37°C and 250 rpm until it reached OD₆₀₀ of 0.5-0.6, and the culture was induced with 1.0 mM arabinose. After 19 hours cultures were harvested by centrifuging at 4,000 g for 10 minutes at 4°C.

Cultures were resuspended in lysis buffer (50mM phosphate buffer pH 7.0, 300 mM NaCl, 10 mM Imidazole, 1 mM MgCl₂, lysozyme, and DNAase) and stored at -80°C for overnight. The bacterial lysates were centrifuged at 45,000 g for 1 hour at 4°C, and the proteins were purified from the supernatants by Nickel-nitrilotriacetic acid chromatography (Qiagen) following the manufacturer's protocol. Protein was further purified using HiTrap Phenyl HP (GE Healthcare) on AKTA protein purification system (GE Healthcare) following the manufacturer's protocol. All variants were dialyzed against 1X phosphate buffered saline (PBS) pH 7.5 prior to biochemical and biophysical characterizations.

2.8 Gel filtration chromatography

Purified proteins in 1X PBS pH 7.5 were analyzed using a Superdex 200 column (GE Healthcare) and eluted with a flow rate of 0.5 ml/min with proteins detected using absorbance at 280 nm. Standard curve was calculated using molecular weight markers kit (Sigma Aldrich).

2.9 Native gel electrophoresis

Purified proteins in 1X PBS pH 7.5 were mixed with 2X sample buffer (62.5 mM Tris-HCl pH 6.8, 25% glycerol, 1% bromophenol blue) loaded on native gel. Samples were run under non denaturing running buffer (25 mM Tris, 192 mM glycine) with 30 mAmp for 45 minutes. The native gel was composed of a separating gel (10% acrylamide pH 8.8) and a stacking gel (3.9% acrylamide pH 6.8).

2.10 Quantum yield determination

Purified iRFP variants were diluted in 1X PBS pH7.5 until the maximum absorbance at any wavelength is less than 0.1. Nile blue dye was diluted in 200 proof ethanol and diluted until the maximum absorbance at any wavelength is less than 0.1. Absorbance spectrum was obtained using Cary spectrophotometer from 205 to 750 nm. Fluorescence was measured using Tecan M1000 plate reader by exciting at 659 nm and emission from 669 to 800nm. Fluorescence area was calculated for both Nile blue and iRFP variants. Quantum yield was calculated using the equation described by Fery-Forgues et al., 1999.

2.11 Extinction coefficient determination

Purified iRFP variants were diluted into 1X PBS pH7.5 until the maximum absorbance at any wavelength was less than 0.1. As described by Shu et al., 2009, I compared the absorbance values for the proteins at the main peak 694 nm with the absorbance value at the 391 nm peak representing free BV whose extinction coefficient is $39,900 \text{ M}^{-1}\text{cm}^{-1}$.

2.12 Equilibrium unfolding

A 6 M stock of guanidinium hydrochloride (GndHCl) was mixed with PBS to make solutions containing a range of GndHCl concentrations (0, 0.5, 1, 1.5, 2, 2.25, 2.5, 2.75, 3, 3.25, 3.5, 4 and 4.5 M). Aliquots of each protein were mixed with 325 μl of each GndHCl solution, incubated at room temperature for 2 hours in dark, and transferred into transparent flat-bottom 96-well plate (Corning) where fluorescence ($\lambda_{\text{ex}} = 690 \text{ nm}$ and $\lambda_{\text{em}} = 700 \text{ to } 800 \text{ nm}$) was measured. The fluorescence signal at 715 nm was used for

comparison. The free energy of folding ($\Delta G_{\text{folding}}$) and concentration of GndHCl that yields half maximal folding was calculated for each protein by fitting the unfolding data using a linear extrapolation method (Pace et al., 2000).

2.13 Mammalian tissue culture and flow cytometry

HeLa cells (American Type Culture Collection) were cultured in Dulbecco modified eagle medium, Lonza, supplemented with 10% fetal bovine serum (Sigma) and 1% penicillin–streptomycin–glutamine (Hyclone) and maintained at 37 °C and 5% CO₂. Cells were plated in 12-well plates at 8×10^4 cells per well. 24 hours after seeding, cells were transiently transfected using JetPrime (Polyplus transfection) according to the manufacturer's protocol with 0.5 ug of plasmid expressing the iRFP variant along with 0.05 ug of pcDNA4-GFP plasmid as transfection control per well. The culture media of transfected cells was replaced with fresh media 16 hours post transfection. Cells were harvested for analysis at 48 hours post transfection by trypsinization (TrypLE, GIBCO Invitrogen). Cells were analyzed with a FACSCanto II flow cytometer (BD, San Jose, CA) to measure fluorescence intensity of iRFP variants in the APC-Cy7 (Allophycocyanin conjugated with cyanine dye) channel (633 nm laser, 780/60 nm emission filter). At least 10,000 cells were recorded in each sample for analysis. For transient transfection experiments, the APC-Cy7 signal changes were monitored within GFP-positive cells (488 nm laser, 530/30 emission filter) to monitor changes within transfected cells. The reported output signal was calculated by normalizing APC-Cy7 intensity by GFP intensity in order to eliminate differences arising from transfection efficiencies.

Chapter 3

IFP tolerance to random fission and rational fusion

Modified and expanded from manuscript published in *ACS Synthetic Biology*, 2015¹

3.1 Abstract

Gene fission can convert monomeric proteins into two-piece catalysts, reporters, and transcription factors for systems and synthetic biology. However, some proteins can be challenging to fragment without disrupting function, such as near-infrared fluorescent protein (IFP). I describe a directed evolution strategy that can overcome this challenge by randomly fragmenting proteins and concomitantly fusing the protein fragments to pairs of proteins or peptides that associate. I used this method to create libraries that express fragmented IFP as fusions to a pair of associating peptides (IAAL-E3 and IAAL-K3) and proteins (CheA and CheY) and screened for fragmented IFP with detectable near-infrared fluorescence. Thirteen novel fragmented IFP were identified, all of which arose from backbone fission proximal to the interdomain linker. Either the IAAL-E3 and IAAL-K3 peptides or CheA and CheY proteins could assist with IFP fragment complementation, although the IAAL-E3 and IAAL-K3 peptides consistently yielded higher fluorescence. These results demonstrate how random gene fission can be coupled to rational gene fusion to create libraries enriched in fragmented proteins with AND gate logic that is dependent upon a protein-protein interaction, and they suggest that these near-infrared fluorescent protein fragments will be suitable as reporters for pairs of promoters and protein-protein interactions within whole animals.

¹Pandey, N. et al., 2015. Combining random gene fission with rational fusion to discover near-infrared fluorescent protein fragments that report on protein-protein interactions. *ACS Synth. Biol.* 4, 615-624.

3.1 Introduction

Component limitations arise in synthetic biology because a restricted number of biological parts can be reliably used to program cellular behaviors (Grunberg et al., 2010; Marcheschi et al., 2013). Gene fragmentation represents a simple strategy to overcome component limitations by converting proteins encoded by individual genes into proteins encoded by two or more gene fragments (Grunberg et al., 2010; Marcheschi et al., 2013). Protein fragments arising from fission display “AND” gate logic when the ORFs encoding the different fragments are placed under distinct transcriptional regulation so that protein fragment complementation only occurs when the promoter driving transcription of the first ORF is active AND the promoter controlling transcription of the second ORF is also active (Shis et al., 2013). Unfortunately, researchers often cannot predict *a priori* where genes (and proteins) can be fragmented into pieces to construct AND gates. In these cases, laboratory evolution can be used to identify which of the many possible fission sites within a protein are least disruptive to function (Ostermeier et al., 1999). This combinatorial approach has the advantage that it samples all of the possible designs for constructing fragmented proteins and provides information on the backbone fission sites that yield fragments with the strongest functional complementation. In addition, comprehensive information on protein tolerance to fission can be used to guide the construction of more complex logic gates such as three-input AND gates (Segall-Shapiro et al., 2014; Cabantous et al., 2013).

Fragmented proteins often display decreased activity compared with their natural counterparts encoded by single polypeptides, which can limit their utility for synthetic biology applications (Burbaum et al., 1991). One way to create fragmented proteins with

improved activity is to introduce non-disruptive backbone fission sites into homologs having enhanced stability, such as a thermophilic homolog of a fragmented mesophilic protein (Nguyen et al., 2008). Highly stable proteins can also be used as starting points for laboratory evolution experiments to increase the fraction of fragmented proteins that retain function within a combinatorial library (Segall-Shapiro et al., 2011), provided that a highly stable homolog is available as a starting point for laboratory evolution. An alternative way to enhance the cooperative function of protein fragments is to fuse protein fragments to a pair of proteins that form a stable complex and promote complementation (Michnick et al., 2007). Assisted protein-fragment complementation has been used to create a variety of genetically-encoded devices whose molecular outputs report on a protein-protein interaction, including devices that generate metabolic (Remy et al., 1999), visual (Rossi et al., 1997; Luker et al., 2004; Hu et al., 2003), transcriptional (Fields et al., 1989; Russ et al., 1999; Karimova et al., 1998), proteolytic (Johnsson et al., 1994; Williams et al., 2009), and antibiotic resistance (Galarneau et al., 2002; Paschon et al., 2005) outputs. The design of fragmented proteins that report on molecular interactions can be arduous (Paschon et al., 2004; Paschon et al., 2005; Choe et al., 2005), and no laboratory evolution method has been described to simplify the discovery of protein fragments which require assistance for complementation.

The Silberg Lab recently described a transposon mutagenesis approach for constructing libraries of vectors that express different fragmented variants of a protein (Segall-Shapiro et al., 2011). Libraries created using this method can be coupled to screens and selections to convert any protein into a genetically-encoded AND gate, provided that one or more pairs of fragments derived from that protein cooperatively

function. Because previous studies have shown that interacting proteins can assist with the complementation of protein fragments (Michnick et al., 2007), I sought to explore whether this library approach could be extended so that it generates each possible two-fragment variant of a protein as fusions to a pair of proteins that associate to form a complex (Figure 3.1). I hypothesized that fragmented proteins mined from this type of library would require associating proteins for maximal activity. I focused my efforts on the near-infrared fluorescent protein (IFP), a mutant of the *Deinococcus radiodurans* bacteriophytochrome BphP which displays spectral properties that are compatible with whole body imaging in animals (Shu et al., 2009). This protein uses BV as a chromophore and folds into a structure containing a knot (Wagner et al., 2005; Wagner et al., 2007). The extent to which IFP can be fragmented to create bimolecular fluorescence complementation assays has not been explored using laboratory evolution, although fragmented IFP have been discovered through rational design which require assistance from a protein-protein interaction for fragment complementation (Filonov et al., 2013; Tchekanda et al., 2014).

To better understand IFP tolerance to fragmentation and the best architectures for building bimolecular near-infrared fluorescence complementation devices, I developed a simple method to construct libraries encoding the different possible fragmented IFP as fusions to pairs of associating peptides (and proteins) and screened for IFP that display near-infrared fluorescence. I examined whether the fluorescent fragmented IFP required assistance from protein-protein interactions for complementation, and I investigated how the biochemical properties of fragmented IFP differed from full-length IFP.

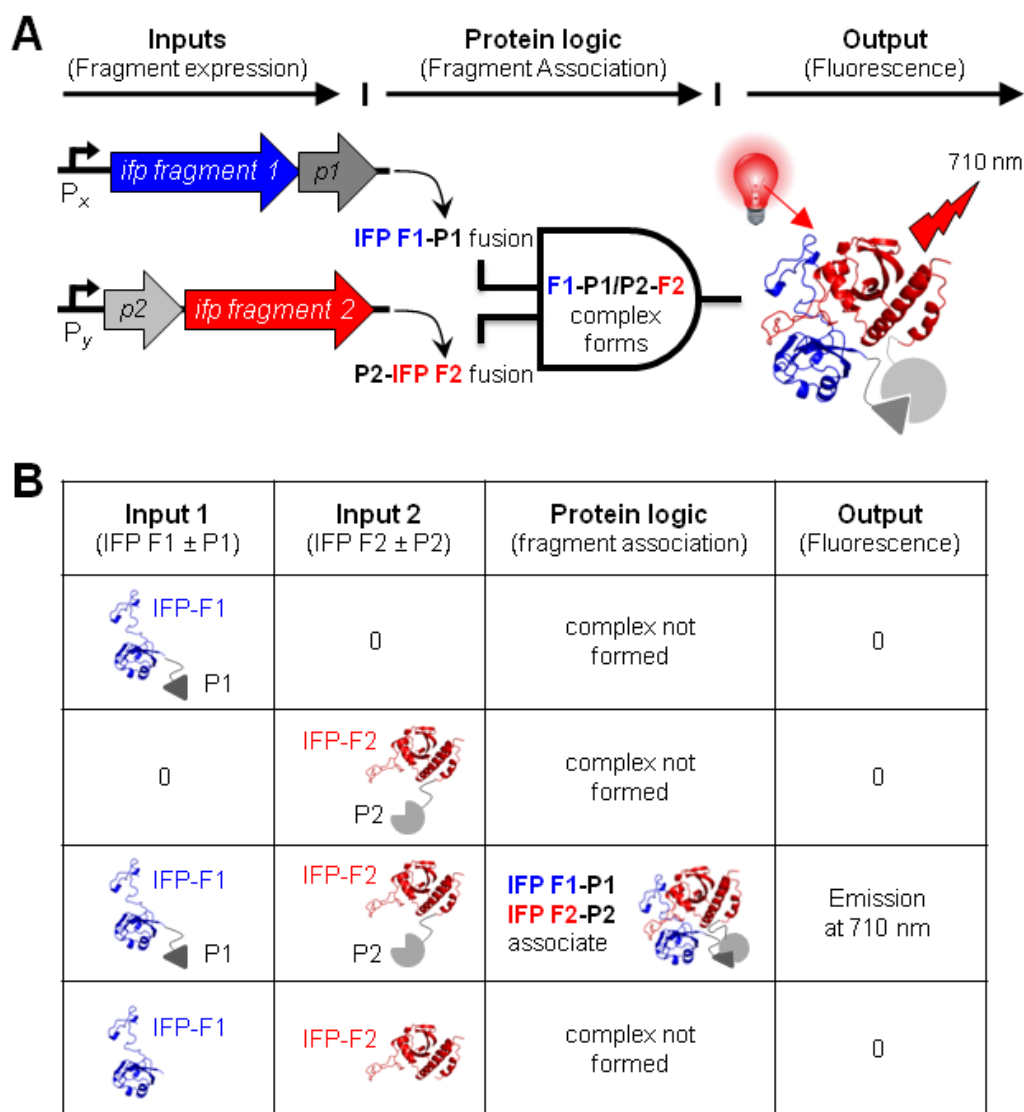


Figure 3.1 Fragmented IFP created by random fission and rational fusion. **(A)** Transposon mutagenesis creates clones that use two promoters (P_x and P_y) to control the expression of IFP fragments (F1 and F2) fused at their termini to a pair of proteins (P1 and P2) that associate. When both fusion proteins (F1-P1 and P2-F2) are expressed, they associate to form a folded two-fragment IFP that binds BV and emits in the near infrared. **(B)** Truth table for a two-fragment protein AND gate that requires fusion to associating proteins for an output.

3.3 Results

3.3.1 Combining IFP fission with fusion to peptides

To establish where IFP can be fragmented into polypeptides that cooperatively fold without assistance, we first created a library of vectors that express fragmented IFP using transposon mutagenesis (Segall-Shapiro et al., 2011) and screened this library for variants that fluoresce like IFP in the near-infrared upon expression within *E. coli*. Screening this library did not identify any two-fragment IFP that retained fluorescence, suggesting that the IFP structure is easily disrupted by fragmentation. Because previous studies have shown that interacting proteins can assist with the complementation of some fragmented proteins (Michnick et al., 2007), I hypothesized that combining random IFP fission with fusion to associating proteins would enrich the library in variants whose fragments associate into an IFP-like structure, bind the BV chromophore, and fluoresce in the near-infrared.

To test the combined effects of random fission and rational fusion on IFP fluorescence, I constructed a library that fuses a pair of peptides (IAAL-E3 and IAAL-K3) designed to associate strongly ($K_D = 70$ nM) to the IFP fragment termini (Litowski et al., 2002). In this EK library, the IAAL-E3 peptide was added to the C-terminus of IFP fragment that precede the backbone fission site, whereas the IAAL-K3 peptide was added to the N-terminus of the IFP fragment that follows the fission site. The IAAL-E3 and IAAL-K3 peptides were chosen for initial library construction because they are small (21 residues each), form a heterodimeric coiled-coil (Litowski et al., 2002), and have been previously used to assist with fragment complementation (Nguyen et al., 2008). Structural studies have shown that the N- and C-termini of IAAL-E3 and IAAL-K3 that

become fused to IFP fragment termini have a physical separation of ~ 30 Å (Lindhout et al., 2004). To minimize IFP structural disruption upon fusion of fragments to termini having this separation, twelve residue glycine-rich linkers were used for library construction.

3.3.2 IFP tolerates fission when fused to peptides

Screening 1,760 clones from the EK library identified multiple colonies with $\lambda_{em} > 700$ nm, and sequencing vectors from these colonies identified seven unique fragmented IFP having their peptide backbone cleaved distal (>25 residues) from the termini. Mapping the locations of these backbone cleavage sites onto the *Deinococcus radiodurans* BphP chromophore-binding domain (Wagner et al., 2005), the protein that was mutated to create IFP (Shu et al., 2009), revealed that all of the fission sites are proximal to the linker that connects the PAS and GAF domains within IFP (Figure 3.2A). These fission sites occur at the end of the second β -sheet within PAS domain (117- and 118-EK), the end of the third β -sheet within PAS domain (129-EK and 131-EK), and the beginning of first helix within GAF domain (140-EK, 142-EK, 144-EK).

3.3.3 Split IFP fluoresce better at 23°C

The finding that peptide fusion was required to discover fragmented IFP with near-infrared fluorescence suggested that these proteins may be marginally stable and display fluorescence that is sensitive to increasing temperature. To test this idea, I compared the near-infrared fluorescence of each variant with full-length IFP at 23 and 37°C.

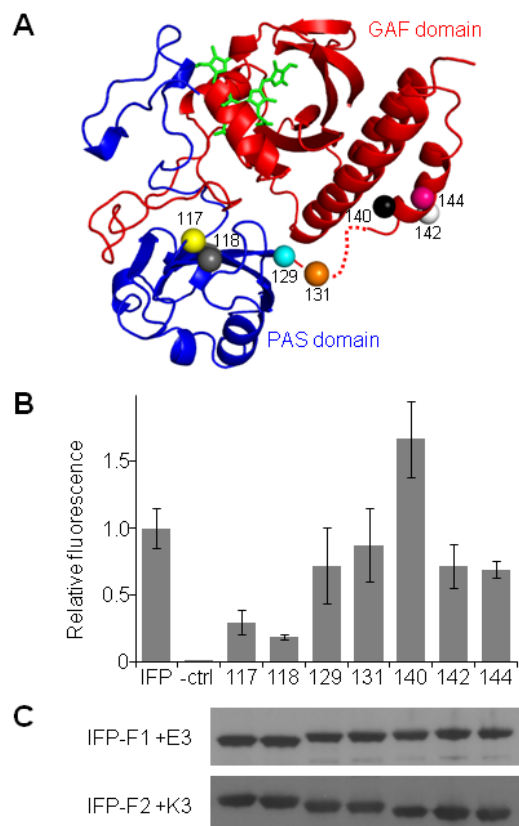


Figure 3.2 Fluorescent fragmented IFP from the EK library. **(A)** For those variants that displayed near infrared fluorescence, the IFP backbone fission sites (spheres) were mapped onto the PAS (blue) and GAF (red) domains of BphP, PDB = 1ZTU,²⁵ using Pymol. Each site represents the last IFP residue in the N-terminal fragment. The domain linker that is disordered in the BphP structure is shown as a dashed line. **(B)** Fluorescence of each fragmented IFP upon expression in *E. coli* is shown relative to that of cells expressing full-length IFP and lacking an IFP (-ctrl). Whole cell fluorescence ($\lambda_{\text{ex}} = 684$ nm; $\lambda_{\text{em}} = 710$ nm) was measured at 37°C and normalized to cell density. The fluorescence intensity obtained with cells expressing each split variant was significantly different from with cells lacking an IFP (two tailed t test; $p < 0.01$). **(C)** Western blot detection of N-terminal IFP fragments fused to IAAL-E3 (IFP-F1 +E3), which have a GST tag and the C-terminal IFP fragments fused to IAAL-K3 (IFP-F2 +K3), which have a HA tag. Error bars represent $\pm 1\sigma$ calculated using three or more replicates.

A comparison of *E. coli* expressing each fragmented IFP revealed similar emission intensity as full-length IFP when fluorescence was analyzed at 23°C (Figure 3.3). In contrast, greater variability was observed when measurements were performed at 37°C (Figure 3.2B), the temperature where IFP biosensors must function to be useful within animal models. Two fragmented IFP displayed <30% of the IFP emission intensity (117-EK and 118-EK), one displayed higher fluorescence (140-EK), and all others displayed intermediate signals.

3.3.4 Split IFP have similar BV dependence

To determine whether the varying fluorescence intensities of the mutants arose because they had peak emission and excitation wavelengths that differed from IFP (Shu et al., 2009), I measured the fluorescence spectrum of each variant. All of the fragmented IFP displayed maximal excitation and emission within 4 nm of intact IFP (Table 3.1). To investigate whether the variability in whole cell fluorescence arose because some fragmented IFP required a different concentration of supplemental BV for maximal fluorescence, I evaluated the dependence of whole cell fluorescence on the amount of BV added to the cells when expressing each fragmented IFP (Figure 3.4). In all cases, the concentration of BV required for half-maximal fluorescence in whole cells (2.3 to 15.5 μM) was lower than the concentration (80 μM) used for screening the libraries. The amounts of BV required for half-maximal fluorescence with fragmented IFP was also similar to the amount (11.2 μM) required for half-maximal fluorescence with IFP. Furthermore, the concentrations required for half maximal fluorescence did not correlate with the relative emission from each fragmented IFP. These findings suggest that the

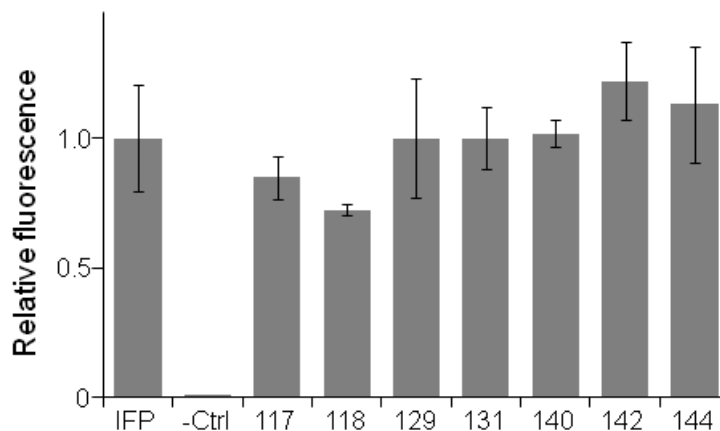


Figure 3.3 Fluorescence of fragmented IFP at 23°C. The fluorescence ($\lambda_{\text{ex}} = 684 \text{ nm}$; $\lambda_{\text{em}} = 710 \text{ nm}$) of *E. coli* expressing fragmented IFP identified in the EK library was measured at 23°C, normalized to cell density, and reported relative to the fluorescence of cells expressing full-length IFP and lacking an IFP (-Ctrl). Error bars represent $\pm 1\sigma$. The fluorescence intensity obtained with cells expressing each split variant was significantly different from with cells lacking an IFP (two tailed t test; $p < 0.005$).

ORF1 (residues)	ORF2 (residues)	+EK		-fusion		+K		+AY	
		λ_{ex} (nm)	λ_{em} (nm)	λ_{ex} (nm)	λ_{em} (nm)	λ_{ex} (nm)	λ_{em} (nm)	λ_{ex} (nm)	λ_{em} (nm)
1-108	108-321	683	709	682	709	---	---	682	712
1-111	111-321	684	712	682	712	---	---	683	712
1-117	117-321	682	711	nd	708	nd	710	683	711
1-118	118-321	683	713	nd	711	nd	715	683	711
1-129	129-321	682	712	682	711	683	709	683	711
1-131	131-321	682	711	682	709	681	712	683	712
1-135	135-321	683	710	682	711	---	---	683	709
1-140	140-321	682	711	683	710	684	709	682	712
1-141	141-321	681	710	684	711	---	---	683	711
1-142	142-321	683	710	683	715	681	709	684	710
1-143	143-321	683	711	683	709	---	---	682	711
1-144	144-321	683	711	682	713	682	714	683	712
1-147	147-321	682	711	683	712	---	---	683	711

Table 3.1 Spectral properties of fragmented IFP. Emission and excitation spectra were measured using *E. coli* expressing each fragmented IFP to determine the wavelengths that yield maximum emission (λ_{em}) and excitation (λ_{ex}) at 37°C. For each variant discovered in the EK and AY libraries, spectral analysis was performed with fragmented IFP fused to IAAL-E3 and IAAL-K3 (+EK), CheA and CheY (+AY), or in the absence of any fusions (-fusion). A subset of the fragmented IFPs were also analyzed with only one fragment fused to the IAAL-K3 peptide (+K). ORF1 represents the IFP residues encoded by the open reading frame that precedes the peptide backbone cleavage site, whereas ORF2 represents the IFP residues encoded by the open reading frame that follows the cleavage site. A sharp excitation spectrum was not detected (nd) with several of the fragmented IFP upon removal of IAAL-E3 or IAAL-E3 and IAAL-K3 due to the weak signal. Cells expressing full-length IFP yielded maximal emission upon excitation at 683 nm and maximal excitation when emission was monitored at 711 nm.

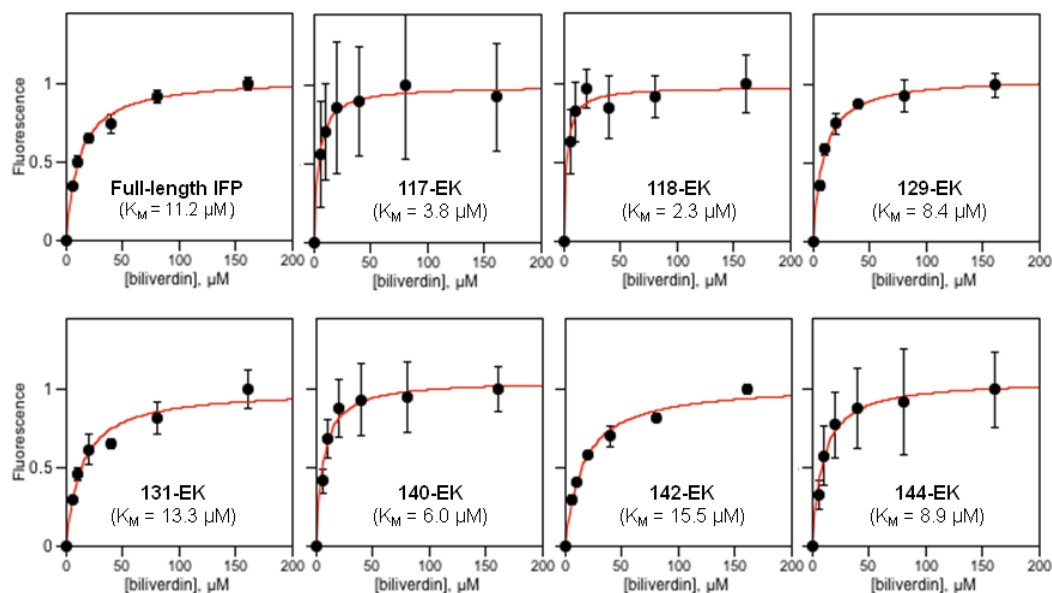


Figure 3.4 Effect of BV concentration on the fluorescence of fragmented IFP fused to IAAL-E3 and IAAL-K3 peptides. The fluorescence ($\lambda_{\text{ex}} = 684 \text{ nm}$; $\lambda_{\text{em}} = 710 \text{ nm}$) of *E. coli* expressing different fragmented IFP discovered in the EK library was measured in cells grown in medium supplemented with varying BV amounts (0, 5, 10, 20, 40, 80, and 160 μM). Fluorescence normalized to the maximum value obtained for each variant was fit to a hyperbolic saturation function to obtain the concentration required for half maximal signal (K_M) within whole cells. Error bars represent $\pm 1\sigma$.

variability in emission does not arise because fragmented IFP differ in the fraction of protein that contains bound BV. To test whether the variation in emission of the IFP variants arose because the different fragment pairs accumulated to varying steady-state levels, I used western blot analysis to evaluate IFP fragment stability (Figure 3.2C). These experiments revealed that the different IFP fragments accumulated to similar extents when expressed from the same promoters and RBS and did not reveal a correlation between near-infrared emission and protein fragment expression. Taken together, these findings suggests that the differences in fluorescence arise because of structural and/or fragment association differences among the split IFP.

3.3.5 Split IFP function as AND gates

The IFP BV chromophore makes a large number of non-covalent interactions with residues within the C-terminal GAF domain that reside within the C-terminal fragments (Figure 1.3). This clustering of BV-interacting residues suggested that fragmented IFP may only require BV association with one fragment to fluoresce. To test this idea, I cloned each pair of IFP gene fragments into vectors that use different inducible promoters (P_{T5} and P_{BAD}) to control expression (Figure 3.5) and examined whether the maximum signal required expression of both N- and C-terminal fragments. Whole cell fluorescence measurements revealed that each of the fragmented IFP displayed the highest fluorescence when *E. coli* harboring these vectors were grown in the presence of both arabinose and IPTG. 140-EK displayed the highest fluorescence as observed with the constructs discovered in the EK library, and similar whole cell fluorescence as cells

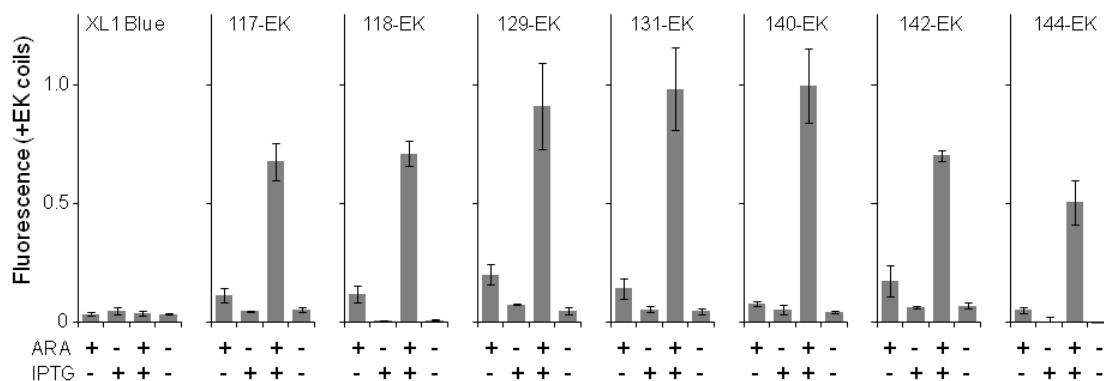


Figure 3.5 Two-input transcriptional regulation of fragmented IFP. The near infrared fluorescence of *E. coli* (XL1-Blue) transformed with vectors that use IPTG- and arabinose-inducible promoters to express the N- and C-terminal fragments of IFP, respectively. The individual and combined effects of 5 mM arabinose (ARA) and 5 mM IPTG on whole cell fluorescence ($\lambda_{\text{ex}} = 684 \text{ nm}$; $\lambda_{\text{em}} = 710 \text{ nm}$) was measured at 37°C, normalized to cell density, and reported relative to the fragmented IFP with the largest signal. Error bars represent $\pm 1\sigma$.

expressing full-length IFP from either inducible promoter (Figure 3.6). In all cases, I detected a low level of near-infrared fluorescence with cells grown only in the presence of arabinose. This trend was interpreted as arising from basal expression of the fragments under P_{T5} promoter regulation, which has been observed previously with the vectors used for the measurements (Mahdavi et al., 2013). These results show that all seven of the fragmented IFP exhibit AND gate genetic logic.

3.3.6 Fragmented IFP require assistance for fluorescence

The paucity of fluorescent proteins discovered when IFP was subjected to fission alone suggested that the fragmented IFP mined out of the EK library might require assistance from the IAAL-E3 and IAAL-K3 coils for maximal fluorescence. This idea was tested by analyzing the fluorescence of cells expressing each fragmented IFP without fusion to these associating peptides. All of the fragmented IFP lacking peptides displayed <10% of the whole cell fluorescence observed with full-length IFP (Figure 3.7A) and lower fluorescence than identical IFP fragments fused to the IAAL-E3 and IAAL-K3 peptides. Among the different fragmented IFPs that fluoresce, peptide removal led to a loss of fluorescence that varied between 7- and 40-fold (Figure 3.7B).

To investigate why peptide removal decreased fluorescence of the fragmented IFP, I analyzed the fluorescence spectra, fluorescence dependence on BV, and steady-state expression of fragments lacking peptide fusions. Fragmented IFP lacking peptides displayed similar excitation and emission maxima as homologous variants fused to the IAAL-E3 and IAAL-K3 peptides (Table 3.1). In contrast, higher concentrations of BV (25.3 to 76.4 μ M) were required for half-maximal fluorescence with IFP fragments

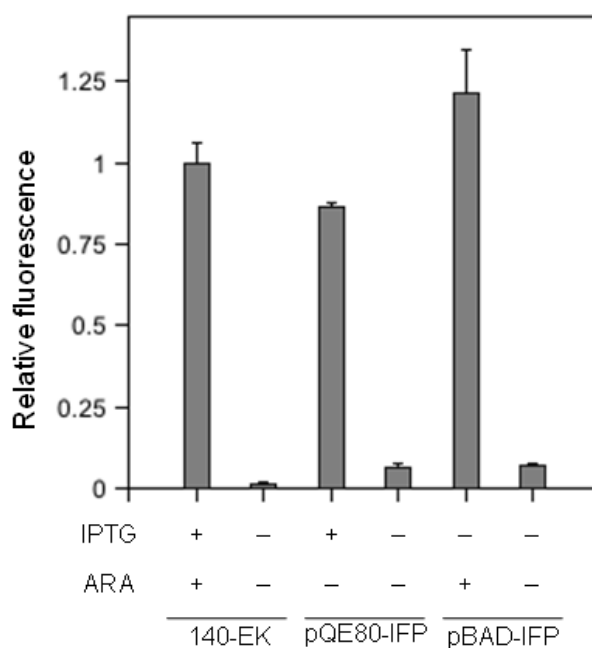


Figure 3.6 140-EK displayed similar fluorescence to full-length IFP from either inducible promoter. The near infrared fluorescence of *E. coli* (XL1-Blue) transformed with vectors that use IPTG- and arabinose-inducible promoters to express the N- and C-terminal fragments of 140 EK respectively and compared to vector that use same promoters to express full-length IFP separately. The individual and combined effects of 5 mM arabinose (ARA) and 5 mM IPTG on whole cell fluorescence ($\lambda_{\text{ex}} = 684 \text{ nm}$; $\lambda_{\text{em}} = 710 \text{ nm}$) was measured at 37°C, normalized to cell density, and reported relative to the fragmented IFP with the largest signal. Error bars represent $\pm 1\sigma$. The fluorescence intensity obtained with cells expressing each split variant or full-length without inducer(s) were significantly different from with cells with inducer(s) (two tailed t test; $p < 0.005$).

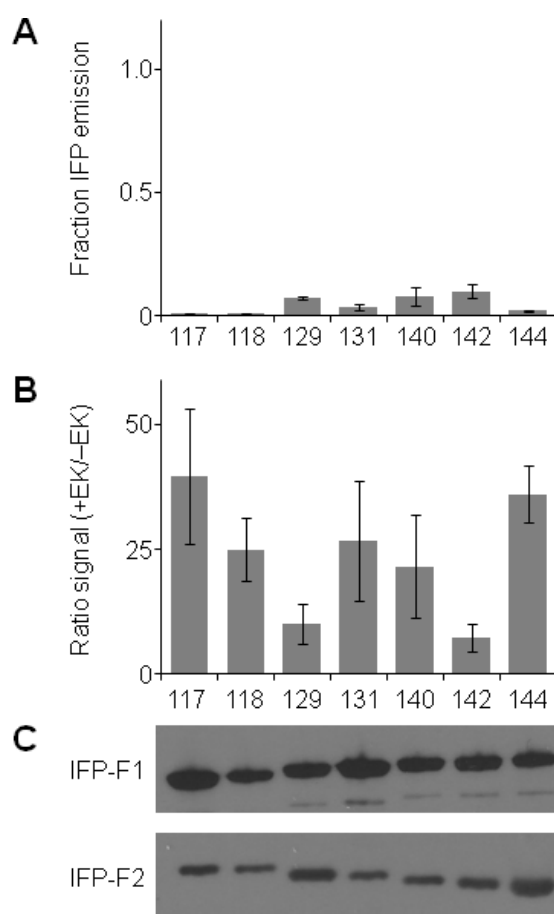


Figure 3.7 IAAL-E3 and IAAL-K3 peptide removal diminishes IFP fragment complementation. **(A)** Whole cell fluorescence ($\lambda_{\text{ex}} = 684 \text{ nm}$; $\lambda_{\text{em}} = 710 \text{ nm}$) of fragmented IFP lacking peptide fusions is shown as the fraction of signal observed with cells expressing full-length IFP. Fluorescence was measured at 37°C and normalized to cell density. **(B)** The ratio of fluorescence measured for two-piece IFP peptide fusions (+EK) to homologous fragmented IFP lacking peptides (−EK). Removal of the EK coils from each split IFP led to a significant change in the fluorescence intensity of each variant (two tailed t test; $p < 0.01$ for all except variant 129 which had a $p < 0.02$). **(C)** Western blot detection of IFP fragments expressed in *E. coli* without IAAL-E3 and IAAL-K3 peptides. N-terminal IFP fragments (IFP-F1) were detected using an antibody against their GST tag, and C-terminal fragments (IFP-F2) were detected using an antibody against their HA tag. Error bars represent $\pm 1\sigma$.

lacking the IAAL-E3 and IAAL-K3 peptides (Figure 3.8). This finding indicates that these variants only displayed a fraction of their maximal whole cell fluorescence under the condition where library screening was performed (80 μ M BV). Western blot analysis revealed that the different IFP fragments lacking peptides accumulated to similar extents (Figure 3.7C) and did not reveal a strong correlation between near-infrared emission and the relative expression of the different fragmented variants lacking fusion to peptides. A comparison of a subset of the fragmented IFP containing IAAL-E3 and IAAL-K3 and homologous fragmented IFP lacking these peptides (Figure 3.9), however, revealed that removal of the peptides from the fragmented IFP altered steady-state levels of each homologous polypeptide fragment. Removal of IAAL-E3 and IAAL-K3 increased accumulation of the N-terminal IFP fragments preceding the peptide backbone cleavage site and decreased accumulation of the C-terminal fragments following the cleavage site. I posited that the decreased expression of C-terminal fragments upon IAAL-K3 removal arose in part because translation initiation of this fragment was decreased by changing the context of the RBS driving expression. This idea was tested by using a thermodynamic model to calculate the relative translation initiation rates from the RBS controlling IFP fragment expression (Salis et al., 2009). All seven of the C-terminal fragments had lower calculated translation initiation rates upon removal of the IAAL-K3 peptide (Figure 3.10). To test this prediction, I created vectors that expressed all of the IFP variants with only one fragment (C-terminal) fused to a peptide (IAAL-K3). Thermodynamic calculations predicted that all of these vectors would initiate translation of the IFP fragments to the same extent as fragmented IFP fused to both IAAL-E3 and IAAL-K3. I found that fragmented IFP fused to only one peptide displayed 1 to 7.5% of the whole

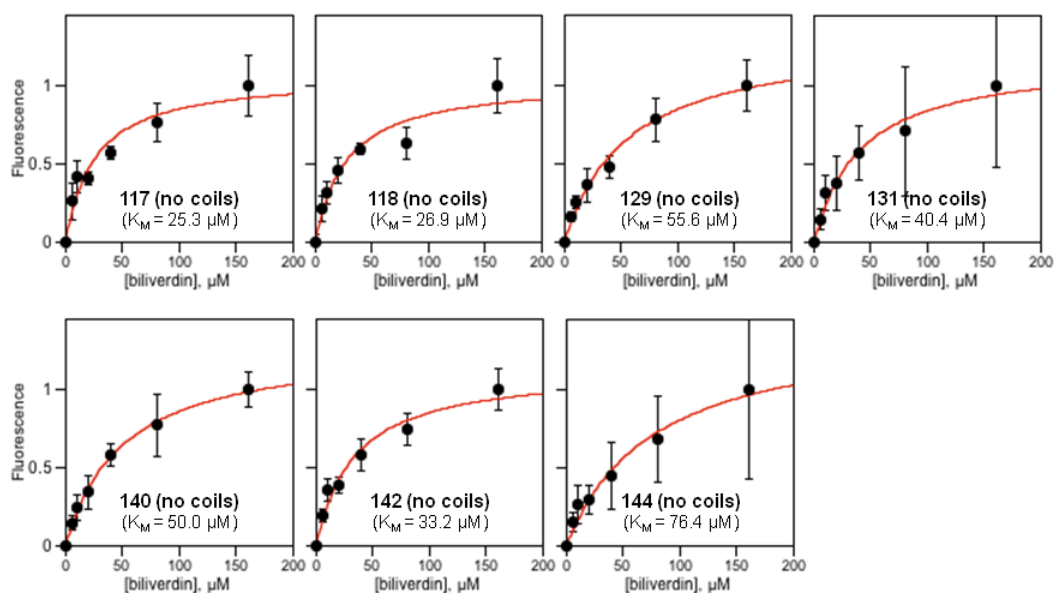


Figure 3.8 Effect of biliverdin concentration on the fluorescence of fragmented IFP lacking fusion to any peptides or proteins. The fluorescence ($\lambda_{\text{ex}} = 684 \text{ nm}$; $\lambda_{\text{em}} = 710 \text{ nm}$) of *E. coli* expressing fragmented IFP that lack fusion to IAAL-E3 and IAAL-K3 peptides was analyzed in medium supplemented with 0, 5, 10, 20, 40, 80, and 160 μM BV. Fluorescence normalized to the maximum value obtained for each variant was fit to a hyperbolic saturation function to obtain the concentration required for half maximal signal (K_M) within whole cells. Error bars represent $\pm 1\sigma$.

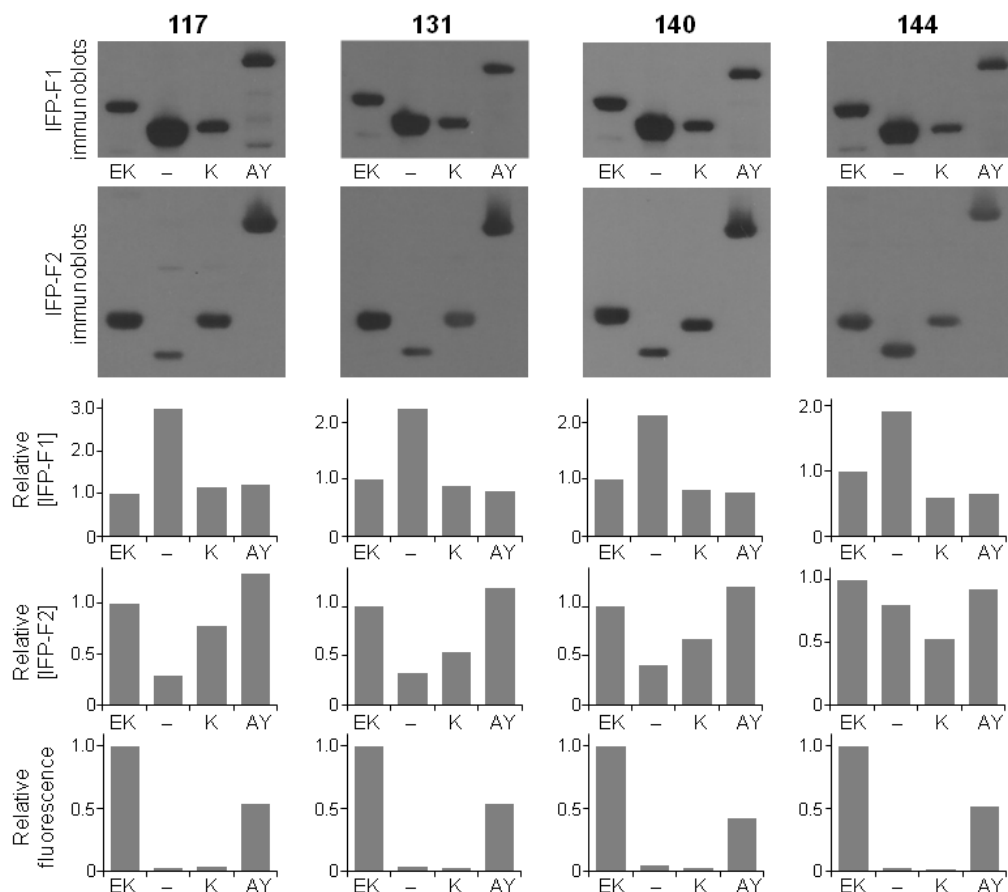


Figure 3.9 Effect of changing protein fusion on the levels of IFP fragments within cells. Western blot analysis was used to analyze how changing the peptides and proteins fused to different fragmented IFP affects the steady-state levels of the fragments under conditions where fluorescence measurements were performed. For four of the fragmented IFP discovered in the EK library, I compared the relative expression of the N- and C-terminal fragments (IFP-F1 and IFP-F2, respectively) of each variant with IAAL-E3 and IAAL-K3 (EK) fused to the different IFP fragments, CheA and CheY (AY) fused to the different IFP fragments, IAAL-K3 (K) fused to only one of the fragments, and no peptides or proteins (-) fused to either fragment. ImageJ was used to quantify the levels of the N- and C-terminal fragments in each blot, which are reported relative to the levels of the two-piece IFP fused to IAAL-E3 and IAAL-K3. For comparison, the bottom panels illustrates the effects of peptide and protein fusion on fluorescence ($\lambda_{\text{ex}} = 684 \text{ nm}$; $\lambda_{\text{em}} = 710 \text{ nm}$).

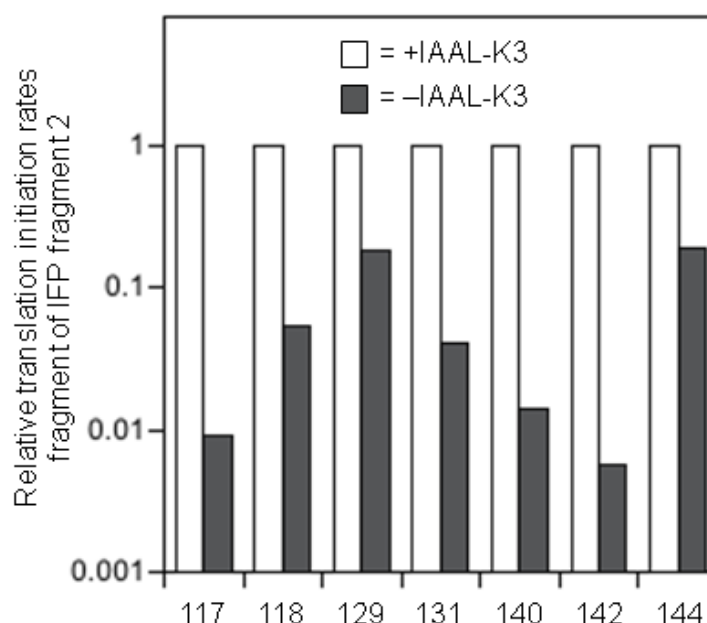


Figure 3.10 Calculated translation initiation rates for C-terminal fragments containing or lacking fusion to IAAL-K3. A thermodynamic model for translation initiation was used to calculate the relative translation initiation rates of the RBS that controlled the expression of IFP fragments containing (open bars) or lacking (closed bars) IAAL-K3 peptides. Translation initiation rates were normalized to the maximum value obtained with IAAL-K3 peptides. All calculations were performed using the DNA sequence that flanked the start codon of the ORFs encoding the IFP fragments, including 50 bases before and after the predicted start codons.

cell fluorescence observed with full-length IFP (Figure 3.11A). This fluorescence emission was similar to or lower than that observed with homologous IFP lacking both peptides (Figure 3.11B). I also found that the excitation and emission spectrum of each variant was comparable to IFP fragments fused to IAAL-E3 and IAAL-K3 (Table 3.1). Western blot analysis did not reveal a strong correlation between near-infrared emission and the relative expression of the different N- and C-terminal fragments (Figure 3.11C). However, I found that the C-terminal fragments fused to the IAAL-K3 peptide accumulated to greater levels than the same IFP fragments lacking the peptide (Figure 3.10) as predicted from RBS strength analysis.

To test whether other interacting proteins could support the complementation of IFP fragments identified in the EK library, I created vectors that expressed each of the fragmented IFP as fusions to *T. maritima* CheA (P2 domain) and CheY. CheA and CheY were chosen because they associate to form a heterodimer (Park et al., 2009) and have been used to assist with fragment complementation of other proteins (Nguyen et al., 2010). I hypothesized that CheA and CheY would be able to substitute for IAAL-E3 and IAAL-K3 without disrupting the fragmented IFP fluorescence because the distance between the termini of CheA and CheY (~ 30 Å) is similar to the distance between the IAAL-E3 and IAAL-K3 termini fused to IFP fragments (Lindhout et al., 2004; Park et al., 2004).

I found that IFP fragments fused to CheA and CheY displayed between 15 and 71% of the whole cell fluorescence observed with full-length IFP (Figure 3.12A). CheA and CheY enhanced IFP fragment complementation to varying extents over IFP fragments lacking fusions to interacting proteins, ranging from 4- to 22-fold (Figure 3.12B). The

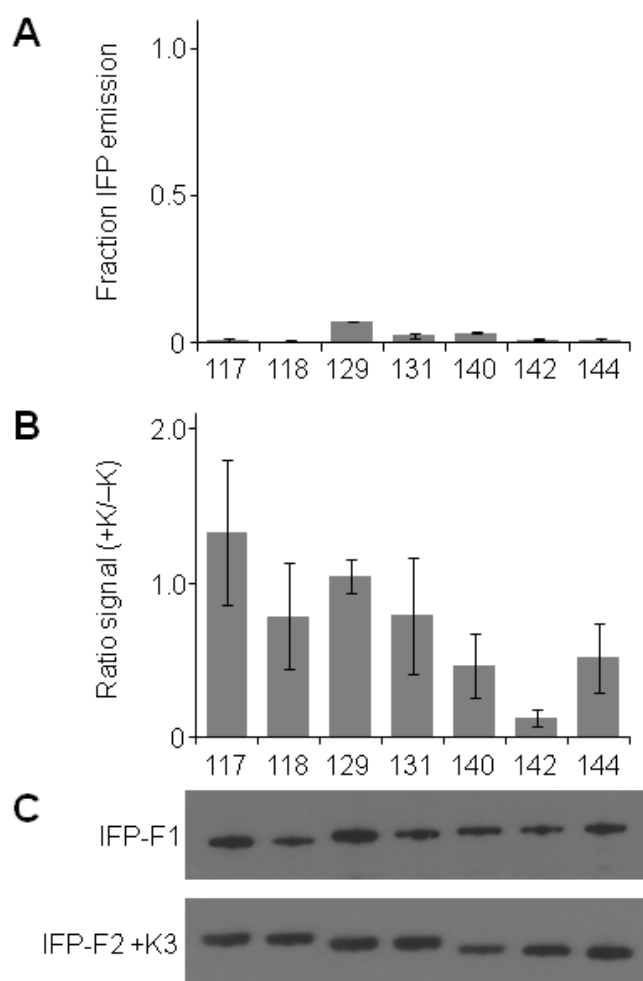


Figure 3.11 Removal of IAAL-E3 decreases IFP fragment complementation. **(A)** Fluorescence of *E. coli* ($\lambda_{\text{ex}} = 684 \text{ nm}$; $\lambda_{\text{em}} = 710 \text{ nm}$) expressing fragmented IFP with their C-terminal fragments fused to IAAL-K3 and their N-terminal fragments lacking a peptide. Fluorescence is reported as the fraction of the signal observed with full-length IFP. **(B)** The ratio of fluorescence measured for fragmented IFP having IAAL-K3 fused at the terminus of one fragment (+K) to fragmented IFP homologs lacking peptide fusions (−K). The fluorescence intensity of split variants containing one K coil was not significantly different from variants lacking both coils, with the exception of variant 144 (two tailed t test; $p < 0.05$). **(C)** Western blot detection of N-terminal IFP fragments (IFP-F1), which have a GST tag, and the C-terminal fragments fused to IAAL-K3 (IFP-F2 + K3), which have an HA tag. Error bars represent $\pm 1\sigma$.

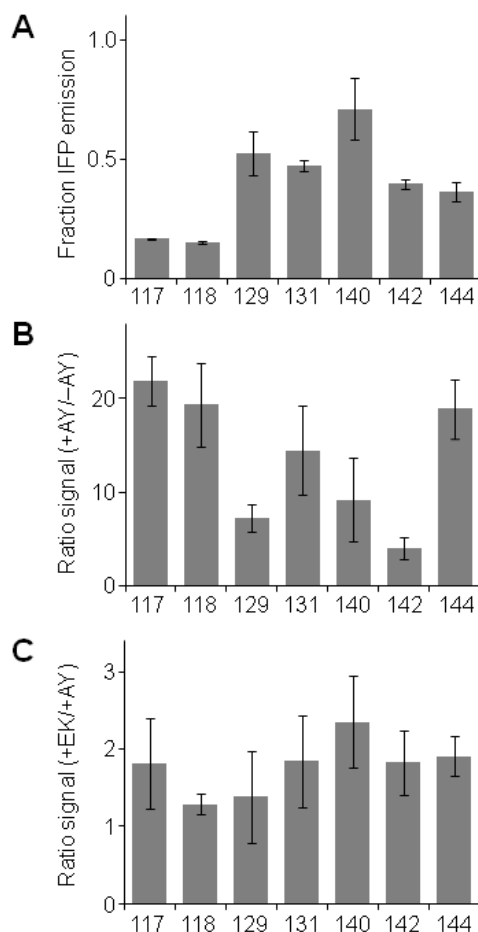


Figure 3.12 CheA and CheY rescue IFP fragment complementation. **(A)** Whole cell fluorescence of *E. coli* expressing each pair of IFP fragments as fusions to CheA and CheY. Fluorescence emission ($\lambda_{\text{ex}} = 684 \text{ nm}$; $\lambda_{\text{em}} = 710 \text{ nm}$) measured at 37°C was normalized to cell density and is shown relative to the signal from cells expressing full-length IFP. **(B)** The ratio of fluorescence measured for fragmented IFP having CheA and CheY fused at their termini (+AY) to homologous IFP fragments lacking CheA and CheY (-AY). The fluorescence intensity obtained with each split variant +AY was significantly different from homologous variants -AY (two tailed t test; $p < 0.005$). **(C)** The ratio of fluorescence for fragmented IFP having IAAL-E3 and IAAL-K3 fused at their termini (+EK) to homologous IFP fragments having CheA and CheY fused at their termini (+AY). Error bars represent $\pm 1\sigma$. Four variants (118, 140, 142, and 144) displayed significantly higher fluorescence +EK compared with +AY (two tailed t test; $p < 0.05$).

relative effects of CheA and CheY on fragmented IFP emission were also similar to the effects of IAAL-E3 and IAAL-K3. IFP fragmented after residues 117 and 144 displayed the largest fluorescence enhancement upon fusion to the different pairs of associating proteins, whereas IFP fragmented after residues 129 and 142 consistently exhibited the smallest enhancement. On average, however, fragmented IFP fused to CheA and CheY displayed lower whole cell fluorescence than the same IFP fragments fused to IAAL-E3 and IAAL-K3 (Figure 3.12C). These differences were not interpreted as arising from changes in spectral properties or fragment expression. Fragmented IFP fused to CheA and CheY displayed spectra with similar excitation and emission maxima as IFP (Table 3.1), and IFP fragments accumulated to a similar level when fused to CheA/CheY and IAAL-E3/IAAL-K3 (Figure 3.9). Instead, these differences are thought to arise from the varying strength of the protein-protein interactions used to assist with IFP fragment complementation. The fragment complementation correlated with the relative affinities of the CheA/CheY ($K_D = 200$ nM) and IAAL-E3/IAAL-K3 ($K_D = 70$ nM) complexes (Litowski et al., 2002; Park et al., 2004).

3.3.7 Combining CheA/CheY fusion with IFP fission

I next sought to determine if combining random IFP fission with fusion to CheA and CheY would yield distinct fragmented IFP from those discovered in the EK library or fragmented IFP with a higher fluorescence. I hypothesized that screening for fragmented IFP from a library created using an alternative protein-protein interaction might identify different fragmented IFP because of differences in fragment expression, stability, and association. To test these ideas, I created a library where CheA and CheY were fused to

the termini of each randomly fragmented IFP. Like the EK library, screening 1,760 clones from this AY library for variants with near-infrared fluorescence yielded multiple variants with $\lambda_{em} > 700$ nm. Sequencing vectors from these cells identified nine IFP that were fragmented distal from the termini (>25 residues). While three of these variants (118-AY, 131-AY, and 140-AY) were identified when screening the EK library, six of the variants had distinct fragmentation sites. Mapping the locations of these sites onto the BphP structure (Wagner et al., 2005) revealed that they cluster near the boundary of the PAS and GAF domain (Figure 3.13A). Two of the backbone fission sites are after the first β -strand of the PAS domain (108-AY and 111-AY), one is within the linker that connects the domains (135-AY), two are within the first helix of the GAF domain (141-AY and 143-AY), and one is after the first helix of the GAF domain (147-AY).

Cells expressing the variants mined out of the AY library displayed only a fraction (16 to 69%) of the fluorescence signal obtained with cells expressing full-length IFP (Figure 3.14). The IFP variants arising from backbone cleavage proximal to the residues linking the PAS and GAF domains showed the highest whole cell fluorescence, similar to that observed with fragmented IFP discovered in the EK library. Analysis of the emission and excitation spectrum of each variant also revealed maximal excitation and emission similar to that of IFP (Table 3.1). I posited that the fluorescence of fragmented IFP discovered in the AY library would depend on fusion to interacting proteins, like the variants discovered from the EK library. To test this idea, I built constructs that expressed each of these fragmented IFP without fusion to any proteins, and measured the near-infrared fluorescence of cells transformed with these vectors. All of these cells expressing

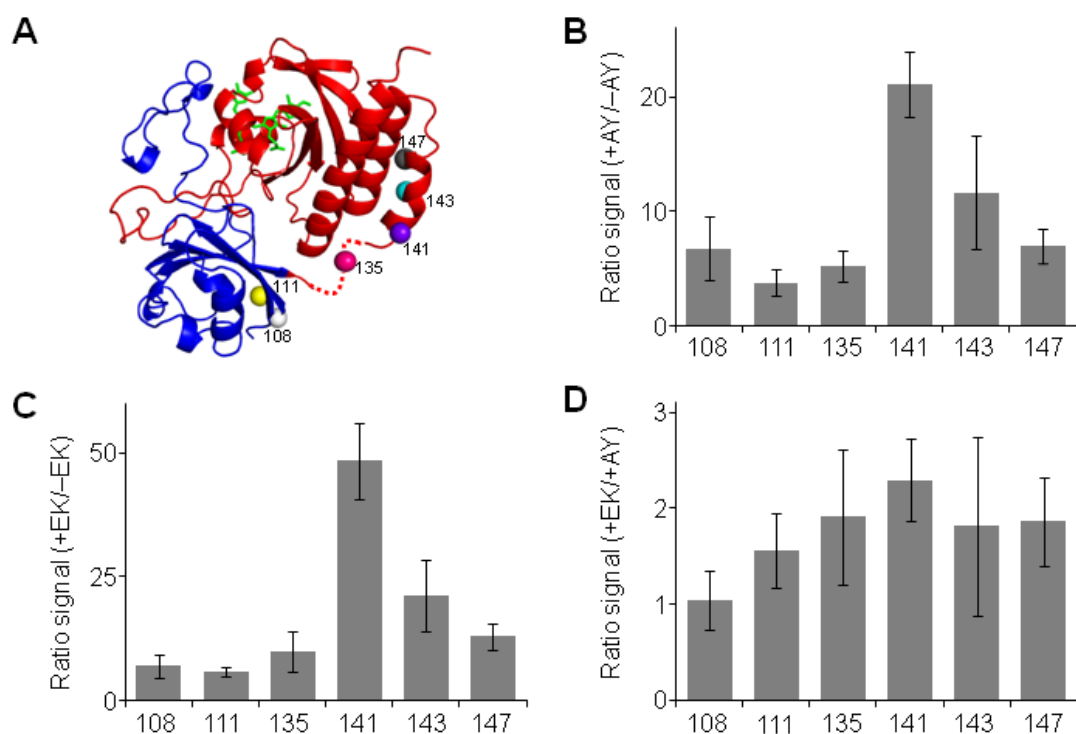


Figure 3.13 Fluorescent fragmented IFP discovered in the AK library. **(A)** The backbone fragmentation sites (spheres) of fragmented IFP discovered in the AY library were mapped onto the PAS (blue) and GAF (red) domains of BphP using Pymol. **(B)** The ratio of fluorescence measured for fragmented IFP having CheA and CheY fused at their termini (+AY) to homologous variants lacking CheA and CheY (-AY). The fluorescence intensity obtained +AY was significantly different from homologous variants -AY (two tailed t test; $p < 0.01$ for all except variant 143 which had a $p < 0.02$). **(C)** The ratio of fluorescence measured for fragmented IFP having IAAL-E3 and IAAL-K3 fused at their termini (+EK) to variants lacking peptides (-EK). The fluorescence intensity obtained +EK was significantly different from -AY (two tailed t test; $p < 0.01$). **(D)** The ratio of fluorescence measured for fragmented IFP having IAAL-E3 and IAAL-K3 fused at their termini (+EK) to variants fused to CheA and CheY (+AY). Three variants (111, 141, and 147) displayed significantly higher fluorescence +EK compared with +AY (two tailed t test; $p < 0.05$). Error bars represent $\pm 1\sigma$.

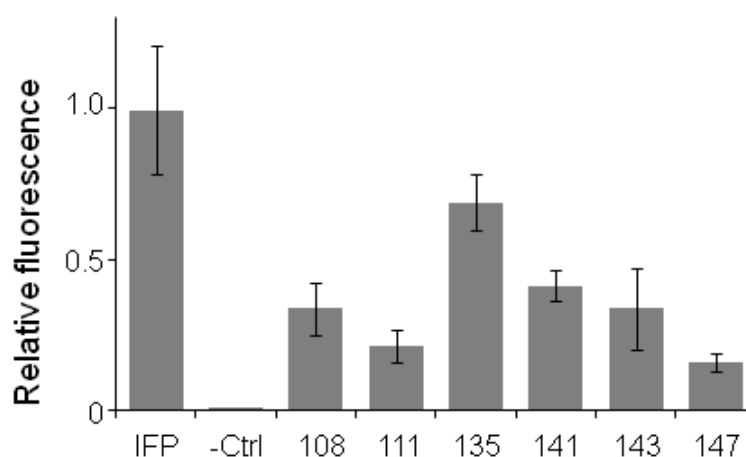


Figure 3.14 Fragmented IFP discovered in the AY library. The fluorescence ($\lambda_{\text{ex}} = 684$ nm; $\lambda_{\text{em}} = 710$ nm) of each fragmented IFP upon expression in *E. coli* at 37°C is shown relative to the fluorescence of cells expressing full-length IFP and lacking an IFP (-Ctrl). Fluorescence is normalized to cell density. Error bars represent $\pm 1\sigma$. The fluorescence intensity obtained with cells expressing each split variant was significantly different from with cells lacking an IFP (two tailed t test; $p < 0.005$ for all except variant 143 for which $p < 0.02$).

fragmented IFP displayed decreased fluorescence upon removal of CheA and CheY (Figure 3.13B), as observed with variants discovered in the EK library. The fluorescence of cells expressing IFP fragments fused to CheA and CheY was 4- to 21-fold higher than cells expressing IFP fragments lacking CheA and CheY fusions. This finding provides additional evidence that random protein fission can be combined with rational fusion to discover near-infrared fluorescence protein fragments that report on protein-protein interactions.

To establish if the variants discovered in the AY library were dependent upon CheA and CheY for maximal fluorescence, I substituted the IAAL-E3 and IAAL-K3 peptides for CheA and CheY, respectively, and analyzed IFP fragment complementation. Substitution of the peptides for CheA and CheY enhanced fragment complementation over that observed with fragmented IFP lacking fusion to any peptides or proteins (Figure 3.13C). As observed with fragmented IFP discovered in the EK library, fluorescence complementation was enhanced to the greatest extent when IFP fragments were fused to the IAAL-E3 and IAAL-K3 peptides (Figure 3.13D). These findings suggest that the affinity of the protein-protein interaction used for library construction contributes to protein-fragment complementation, because the higher affinity interaction (IAAL-E3/IAAL-K3) consistently yielded stronger complementation with the fragmented IFP.

3.4 Discussion

The results described in Chapter 3 show how combining random gene fission with fusion to ORFs encoding associating proteins can aid with the discovery of protein fragments that cooperatively function (fluorescence) and display two-input AND gate logic (two-promoter regulation). The EK and AY libraries yielded fragmented IFP that displayed near-infrared fluorescence, while fluorescent variants were not discovered in the library encoding IFP that had been subjected to random fission alone. The fragmented IFP discovered in the EK and AY libraries provide positional information on the combined effects of fission and fusion on IFPs retention of near-infrared fluorescence. All of the IFP backbone fission sites discovered using laboratory evolution were distal from the BV-binding site, suggesting that IFP fission and fusion near this chromophore binding site is disruptive to fragment complementation. The backbone fragmentation sites identified in this study (108, 111, 117, 118, 129, 131, 135, 140, 141, 142, 143, 144, and 147) were distinct from those evaluated in previous studies that used rational design to sample a smaller number of backbone cleavage sites within IFP (Tchekanda et al., 2014) and iRFP (Filonov et al., 2013), an IFP homolog that displays higher affinity for biliverdin. These studies each identified backbone cleavage sites within the interdomain linker that can be fragmented to build split protein assays that report on protein-protein interactions. The sites clustered proximal to the best site (132) identified in the IFP study and within the secondary structures (α -helixes, β -sheet, and loops) adjacent to this domain linker. Whether or not these new split IFP have any advantages or disadvantages over the rationally designed IFP/iRFP fragment complementation assays will require

additional experiments that directly compare their reversibility, self-association, and signal within cellular models using different pairs of interacting proteins.

The decreased near infrared fluorescence consistently observed when the associating peptides and proteins were removed from the thirteen fragmented IFP is interpreted as arising because each of these IFPs require assistance from interacting proteins for fragment complementation. Evidence for this idea comes from the finding that IAAL-E3 removal diminished fluorescence complementation dramatically (10- to 70-fold) while only having minor effects on IFP fragment levels. Among the three libraries screened for protein fragments displaying near-infrared fluorescence, only the IAAL-E3/IAAL-K3 and CheA/CheY libraries yielded fragmented IFP with strong signals. The dependence on associating peptides (or proteins) for discovery of fragmented IFP can be contrasted with past fission studies. Using laboratory evolution, larger numbers of non-disruptive backbone fragmentation sites have been discovered within *Thermotoga neapolitana* adenylate kinase (Segall-Shapiro et al., 2011), *E. coli* methionyl tRNA synthetases (Mahdavi et al., 2013), and T7 phage RNA polymerase (Segall-Shapiro et al., 2014). I believe paucity of fragmented IFP that retain strong near-infrared fluorescence without assistance from interacting proteins could arise because the IFP screen could only reliably identify fragmented proteins with near-native function, while previous studies used assays that could detect a fraction of parental protein activity. In addition, IFP fragmentation could decrease its affinity for BV, which is required for fluorescence, and the stability of IFP under the screening conditions. Support for the latter idea comes from the observation that full-length IFP displays almost 4-fold lower whole cell fluorescence when expressed at 37°C compared with 23°C (Figure 3.15).

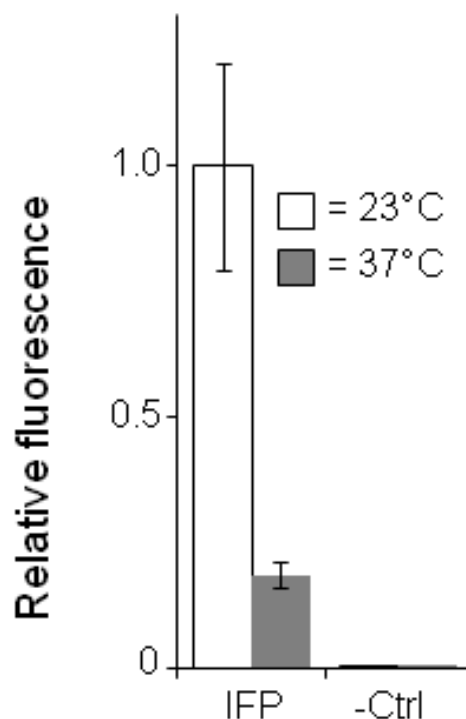


Figure 3.15 IFP displayed increased fluorescence at 23°C. The fluorescence ($\lambda_{\text{ex}} = 684$ nm; $\lambda_{\text{em}} = 710$ nm) of IFP upon expression in *E. coli* at 23°C and 37°C is shown relative to the fluorescence of cells lacking an IFP (-Ctrl). Fluorescence is normalized to cell density. Error bars represent $\pm 1\sigma$. The fluorescence intensity obtained with cells expressing IFP was significantly different from with cells lacking an IFP (two tailed t test; $p < 0.005$) and significantly different IFP fluorescence at 23°C vs. 37°C (two tailed t test; $p < 0.005$).

The transposon mutagenesis approach used to randomly fragment IFP and fuse the fragments to interacting proteins differs from a combinatorial protein engineering method previously described for this application, which uses incremental truncation for random fission (Ostermeier et al., 1999) and blunt ligations for rational fusion (Paschon et al., 2005). In contrast to transposon mutagenesis (Segall-Shapiro et al., 2011), incremental truncation requires the creation of two vector libraries: one that encodes the different N-terminal protein fragments and a second that encodes the C-terminal protein fragments (Paschon et al., 2004). Genes encoding a pair of interacting proteins are then subcloned into each of these libraries using blunt ligation, and the two vector ensembles are cotransformed into cells and screened or selected for pairs of complementing fragments. I expect that transposon mutagenesis will be easier for many synthetic biology labs to implement, because transposon-mediated gene fission does not require time-dependent sampling like incremental truncation (Paschon et al., 2004). In addition, gene fragments created using transposon mutagenesis can be fused to any associating proteins by simply creating an insert encoding a pair of interacting proteins (Figure 2.2) and cloning that insert into the NotI restriction site created by transposon insertion (Figure 2.1), rather than performing multiple low efficiency blunt ligations, which is required with libraries created using incremental truncation. The sequence diversity created by transposon mutagenesis [= protein length - 1] is also more constrained than that created by incremental truncation [= (protein length - 1) · (protein length - 1)]. This occurs because transposon mutagenesis cleaves a gene into two fragments within the context of a single vector by inserting a unique NotI restriction site at different locations. In contrast, incremental truncation generates two vector libraries encoding the different types of

proteins fragments, and screens all possible N-terminal fragments with all possible C-terminal fragments by crossing the two libraries (Paschon et al., 2005). Whether or not the more limited diversity created by transposon mutagenesis is more or less advantageous than that created by incremental truncation will require further experiments. The major advantage of creating limited diversity is that additional targeted diversity can be included in combinatorial libraries, *e.g.*, within the linker region, and more thoroughly sampled using screens whose throughput is limited.

Chapter 4

iRFP tolerance to random circular permutation

4.1 Abstract

Near infrared fluorescent proteins have been generated by mutating bacterial phytochromes, which contain a topological knot. To explore how rearrangements in the sequence of the peptides that interlace into a knot affect iRFP folding, I subjected this phytochrome to random circular permutation using transposase mutagenesis and screened for variants that displayed near infrared fluorescence. I identified 27 circularly permuted iRFP that display biliverdin-dependent fluorescence in *E. coli*. The variants with the brightest whole cell fluorescence initiated translation at residues near the BphP domain linker and knot tails, although fluorescent variants were discovered that initiate translation within the PAS and GAF domains. Permuted iRFP retained sufficient biliverdin affinity to fluoresce in tissue culture without the addition of biliverdin, and one variant displayed enhanced protein-normalized fluorescence in bacteria and tissue culture. This variant displayed a similar quantum yield as iRFP, although it exhibited enhanced resistance to chemical denaturation suggesting that the increased signal arises from more efficient knotted protein maturation. These results show how the contact order of a knotted bacteriophytochrome can be altered without disrupting folding, chromophore binding, and fluorescence. These permuted iRFP should be useful for engineering near-infrared biosensors through domain insertion, and they should be useful for guiding the construction of circularly permuted phytochromes that display light dependent conformational switching.

4.2 Introduction

The first description of a knotted protein was controversial because of the entropic penalty that is associated with interlacing a polypeptide chain (Liang et al., 1995; Mansfield et al., 1994). Since that time, over one thousand knotted structures have been deposited in the protein data bank (Jamroz et al., 2010), suggesting that this structural feature is present in many macromolecules that support cellular reactions (Yeates et al., 2007, Sulkowska et al., 2013, Viranu et al., 2006). Unfortunately, our understanding of the benefits and challenges associated with the design and evolution of different knotted topologies remains limited. A recent experimental study provided evidence that some knotted proteins spontaneously fold, albeit with greatly reduced rates compared with the rates observed in the presence of molecular chaperones (Mallam et al., 2012). In addition, knots are thought to provide extra mechanical stability to proteins (Sulkowska et al., 2008). However, this additional free energy is thought to come at the expense of decreased folding rates, since introduction of a knot into an unknotted protein decreased the folding rate by over an order of magnitude (King et al., 2010). While this study demonstrated that knotted proteins can arise from an unknotted topology, our understanding of knotted protein tolerance to sequence changes that alter protein contact order (circular permutation and domain insertion) remains limited.

One of the simplest ways to study a protein's tolerance to changes in contact order is to create circularly permuted variants in which the original NH₂ and COOH termini are connected through a peptide linker and new termini are created elsewhere through backbone fission (Yu et al., 2011). Previous studies have shown that systematic circular permutation can be used to identify essential folding elements in proteins (Iwakura et al.,

2000). This approach has also been used to create proteins with improved activity (Yu et al., 2010; Daugherty et al., 2013; Sheng et al., 2014), altered ligand binding (Dai et al., 2014; Maatta et al., 2008; Stephen et al., 2012; Gauntas et al., 2012) and enhanced stability (Whitehead et al., 2009). A method was recently described for building permuted protein libraries called PERMutation Using Transposase Engineering (PERMUTE), which simplifies the construction of circularly permuted protein libraries for use in laboratory evolution experiments (Mehta et al., 2012). Application of PERMUTE to a thermophilic adenylate kinase revealed multiple structure-function correlations (Jones et al., 2016). Protein tolerance to circular permutation correlated with the distance between the active site and the location of the new termini. Additionally, the functional variants were enriched in structures with NH₂ and COOH termini at sites that display high sequence variability in a multiple sequence alignment and high positional deviation in superpositions of structures. While these findings suggest that structural correlations may exist in other proteins, it remains unclear how they will relate to proteins whose native structures contain knots.

To better understand knotted protein tolerance to changes in contact order, I subjected a phytochrome to random circular permutation and screened for variants that display parent-like function. Phytochromes are photoreceptors found in many taxa (bacteria, fungi, algae and plants) that use light-dependent conformational changes to alter the activity of signaling domains that affect growth and development (Rockwell et al., 2006; Burgie et al., 2014). In bacteriophytochrome, the photosensory core is made up of fused PAS and GAF domains that interlace into a knot, which use the photochemical isomerization of a covalently bound tetrapyrrole to interconvert between two

conformational states upon exposure to red and far-red light (Wagner et al., 2005; Wagner et al., 2007). I chose the knotted BphP photosensory core as a model system because mutants of the photosensory core have been described that exhibit strong fluorescence upon folding, such as near infrared fluorescent protein (iRFP), which can be readily monitored using measurements of whole cell fluorescence in *E. coli* (Pandey et al., 2015). In addition, a better understanding of BphP tolerance to circular permutation is currently needed to understand the extent to which knotted members of this protein family can be used to create near-infrared biosensors or light-dependent switches.

4.3 Results

4.3.1 Generation of permuted iRFP

I created circularly permuted variants of iRFP (Filonov et al., 2011), a mutant of the photosensory core from *Rhodopseudomonas palustris* BphP2 whose native structure displays near infrared fluorescence upon BV binding. iRFP was chosen for these experiments because it displays enhanced stability compared with the other bacteriophytochromes (Filonov et al., 2011), a property that has been implicated as increasing protein tolerance to different classes of mutational lesions (Bloom et al., 2005). PERMutation Using Transposase Engineering (PERMUTE) was used to construct a library encoding circularly permuted variants of iRFP (Mehta et al., 2012). With PERMUTE, an artificial transposon containing all of the attributes of a vector (called a permuteposon) is randomly inserted into a second vector harboring the gene of interest, and a multistep procedure is used to convert the product of this reaction into a library (Figure 2.3A). To allow for regulated protein expression in PERMUTE libraries, I synthesized a new transposon, referred to as a permuteposon P4 (Figure 2.3C) and randomly inserted P4 into a circular iRFP gene to create a library. Restriction digestion analysis on the final library showed that the library contains an ensemble of vectors (Figure 4.1).

4.3.2 Screening the library identified multiple variants

My vector library was transformed into *E. coli*, and individual colonies were screened at either 23 or 37°C for near infrared fluorescence ($\lambda_{\text{ex}} = 690 \text{ nm}$; $\lambda_{\text{em}} = 715 \text{ nm}$). Two

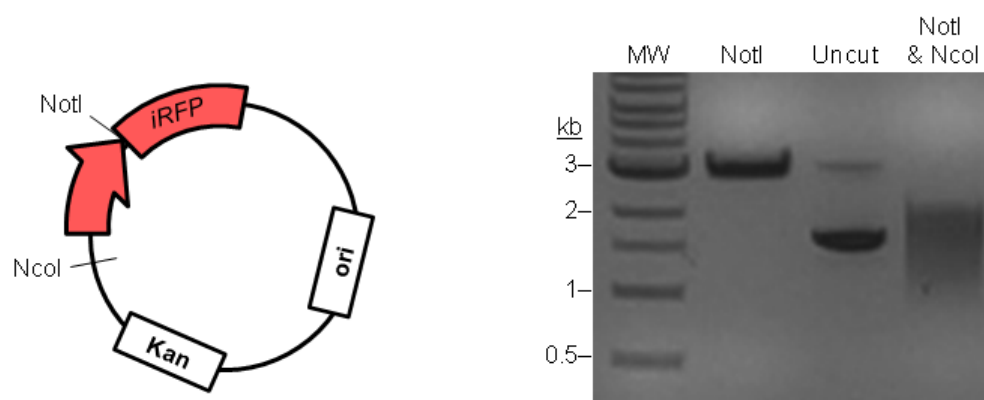


Figure 4.1 Quality assessment of the circularly permuted iRFP library. When digested with NotI, the final library yields a single band that corresponds to the molecular weight of P4 (1911 bp) and the iRFP gene (993 bp). When digested with NotI and NcoI, a smear is observed, consistent with the library containing an ensemble of permuted iRFP.

temperatures were used for screening to increase the likelihood of identifying marginally stable variants. Screening a total of 1760 colonies identified 38 variants with emission $>5\sigma$ over that observed with cells alone. Sequencing vectors from these colonies revealed 27 unique in frame variants that retained near infrared fluorescence characteristic of BV-bound iRFP (Table 4.1). To visualize how these permuted proteins differ in primary structure from iRFP, the first iRFP residue that is translated in each permuted variant was mapped onto the structure of the *Deinococcus radiodurans* BphP chromophore-binding domain (Figure 4.2), a bacteriophytochrome that displays 39% identity with iRFP. This analysis revealed the highest density of variants arising from backbone fission proximal to the domain linker and near the termini of the protein.

4.3.3 Permuted variants display low contact order

To determine how permutation tolerance relates to iRFP folding, the contact order (CO) of permuted protein variants was evaluated. Previous studies have shown that contact order is inversely correlated with protein folding rates (Plaxco et al., 1998), suggesting that knotted proteins might display a low tolerance to increases in contact order. This analysis revealed that fluorescent variants have a range of CO values, including values that are higher and lower than that observed with native iRFP. The highest density of permuted variants in the primary structure had new termini that clustered in a region that yielded CO values that are lower than iRFP (Figure 4.2).

Variant	23°C			37°C			pBAD		
	λ_{ex} (nm)	λ_{em} (nm)	σ	λ_{ex} (nm)	λ_{em} (nm)	σ	λ_{ex} (nm)	λ_{em} (nm)	σ
8	696	715	284	696	715	1452	693	715	1075
11	692	712	101	692	714	330	694	714	1171
12	688	717	43	693	715	242	693	714	1135
15	nd	712	12	nd	715	19	693	713	94
52	692	717	35	692	716	43	nd	713	0
59	nd	715	4	nd	715	7	693	713	20
101	690	714	39	692	714	261	693	714	250
102	686	714	45	692	713	271	692	714	316
119	700	717	58	693	716	114	693	715	298
122	nd	nd	0	692	713	51	693	715	627
123	699	719	14	694	716	105	692	715	747
127	694	714	134	692	714	468	692	715	734
129	nd	nd	0	693	714	45	694	715	595
133	694	714	171	693	715	616	693	715	770
134	693	714	143	692	714	499	693	715	790
136	694	715	214	693	715	796	693	714	513
142	692	716	63	693	717	294	694	715	435
145	nd	718	16	694	720	21	693	716	56
148	nd	720	12	nd	nd	3	693	714	454
182	694	712	27	693	718	44	nd	710	0
189	687	717	10	692	718	15	692	715	214
198	690	715	10	690	713	90	696	713	0
265	nd	712	6	694	714	13	692	716	0
291	nd	711	5	nd	nd	0	nd	717	3
312	693	713	155	692	715	291	693	715	638
313	693	716	257	693	714	360	693	714	963
314	694	713	263	694	714	228	693	714	1231

Table 4.1 Spectral properties of circularly permuted iRFP. Excitation and emission spectra were measured using *E. coli* expressing each permuted iRFP to determine the wavelengths that yield excitation (λ_{ex}) and maximum emission (λ_{em}) at 23°C, 37°C and pBAD vector. Variant represents the first residue of permuted iRFP that belongs to wild-type iRFP sequence. σ represents the signal of the permuted iRFP compared to vector lacking iRFP gene. A sharp excitation spectrum was not detected (nd) with several of the permuted iRFP due to the weak signal. Cells expressing full-length iRFP yielded maximal emission upon excitation at 694 nm and maximal excitation when emission was monitored at 715 nm.

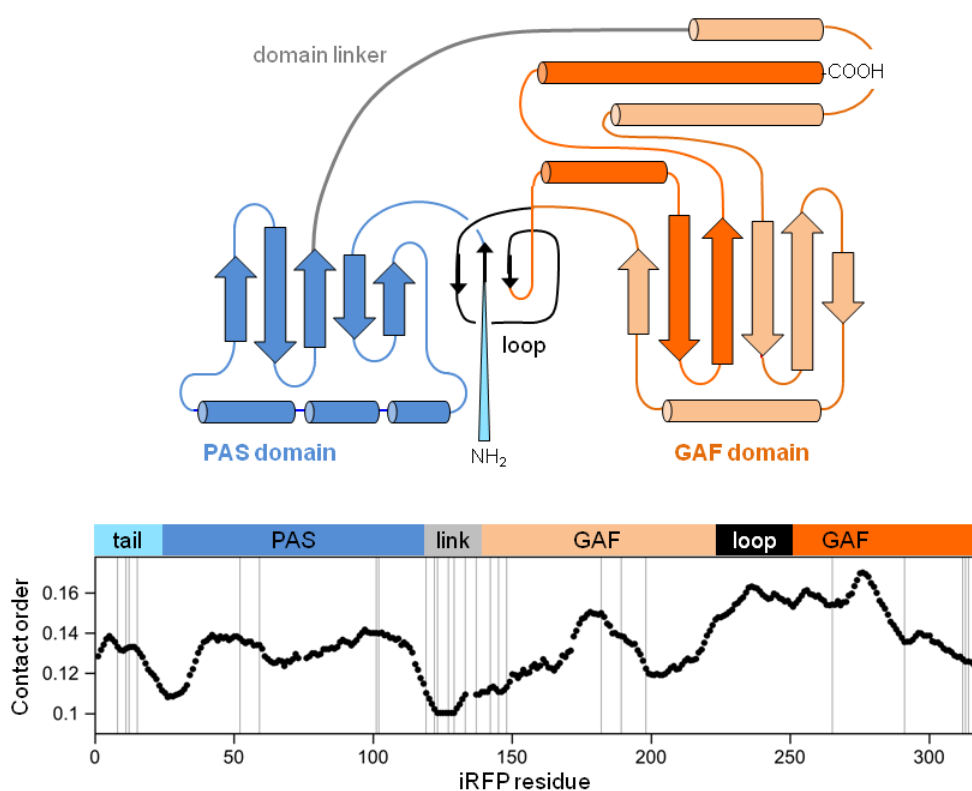


Figure 4.2 Mapping of the iRFP tolerance to permutation and contact order. In iRFP, a knot is formed when the first 33 residues (NH₂ tail) of the PAS domain pass through a loop created by residues from the GAF domain. Relationship between domain structure, contact order, and the location of the new termini created by permutation. The fission sites in fluorescent cp-iRFP are mapped (with lines) onto BphP (PBD=1ZTU) to relate how their location relate to the NH₂ tail (cyan), PAS domain (blue), domain linker (grey), loop from the GAF domain (black) and remainder of the GAF domain (orange). Each line represents the first residue in the circularly permuted iRFP.

4.3.4 A majority of permuted iRFP fluoresce better at 37°C

Distinct sets of circularly permuted iRFP were discovered at the different temperatures where screening was performed (23 and 37°C), suggesting that some of the permuted variants may display fluorescence that varies with temperature. To test this hypothesis, I compared the emission and excitation spectra of each permuted variant with those of native iRFP. Initial spectroscopic measurements at 23°C revealed that my permuted iRFP display similar spectral shapes but a range of peak intensities (Table 4.1). In addition, five variants yielded whole cell emission that was greater than that observed with iRFP, three variants yielded intensities that were comparable to iRFP, and the remaining nineteen variants showed lower emission that was statistically significant (Figure 4.3A). When measurements were performed at 37°C, most variants displayed a lower signal, with only three variants yielded a signal that matched or exceeded the intensity as iRFP. To determine which permuted proteins, similar to iRFP, display an enhanced emission signal as temperature is raised from 23 to 37°C, I calculated the ratio of fluorescence emission for each variant expressed in cells grown at these different temperatures (Figure 4.3B). This analysis revealed that 52% of the circularly permuted iRFP display higher emission at 37°C compared with 23°C like iRFP, 37% have similar fluorescence ratios as iRFP, and 11% display decreased emission as temperature is increased. Among the variants that displayed a decreased signal at 37°C (i.e., fluorescence emission ratios ≤ 1), 80% were discovered in my screen performed at that 23°C.

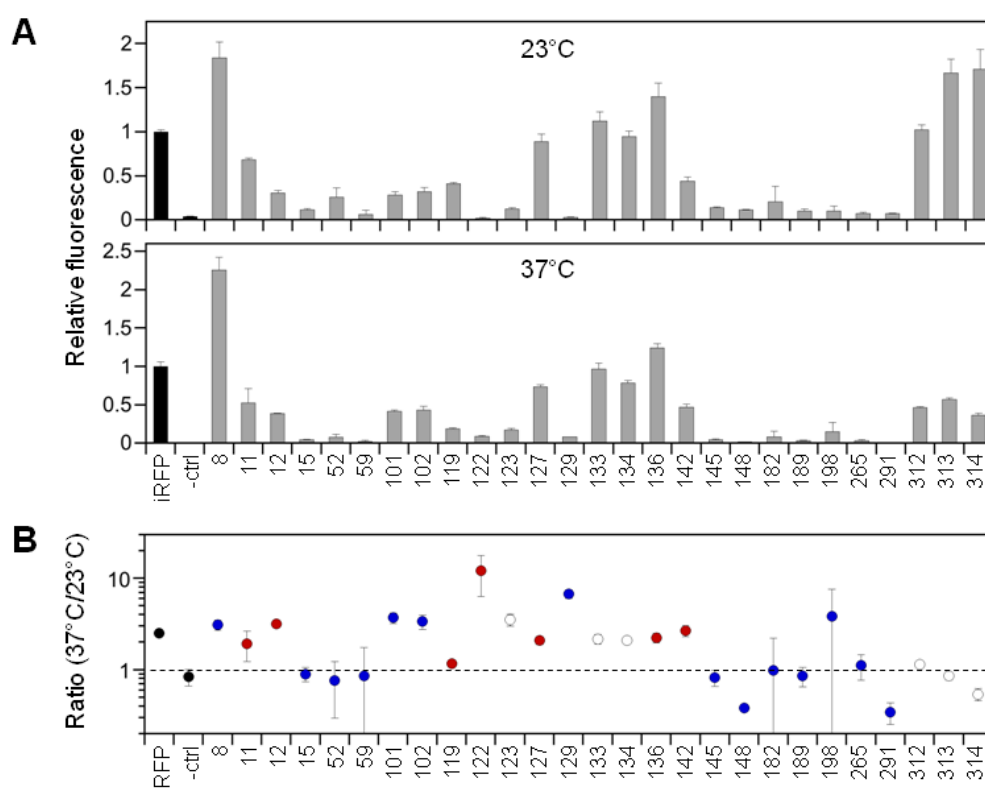


Figure 4.3 Comparison of circularly permuted iRFP variants fluorescence at 23°C and 37°C. **(A)** The fluorescence of cells expressing each variant was measured at 23°C and 37°C ($\lambda_{\text{ex}} = 690 \text{ nm}$; $\lambda_{\text{em}} = 715 \text{ nm}$), normalized to cell density, and compared to the fluorescence of cells expressing parental iRFP and cells lacking iRFP (-ctrl). All circularly permuted iRFP except variants 59, 182, 198 and 291 had significantly higher signal than background at 37°C, and all variants except 59, 122, 129, 182 and 198 had signals that were significantly higher than background at 23°C (two-tailed t test; $p < 0.05$). **(B)** The ratio of the fluorescence signals acquired at 37°C to the signals observed at 23°C. Blue and red symbols represent the variants identified in screens at 23°C and 37°C, respectively, while white symbols represent variant discovered at both temperatures. Error bars represent $\pm 1\sigma$ calculated using three or more independent measurements.

4.3.5 Permuted iRFP have similar BV dependence as iRFP

The variability observed with circularly permuted protein fluorescence intensity could arise because permutation alters protein expression, folding, cofactor binding, and/or stability and degradation. In a previous study, I found that backbone fission increased the concentration of BV required for fluorescence in a related bacteriophytochrome (Pandey et al., 2015), suggesting that backbone fission arising from permutation could have a similar effect on iRFP. To test whether BV affinity is affected by permutation, I evaluated whether near infrared fluorescence varied with the amount of BV added to the cells that are expressing each circularly permuted iRFP (Figure 4.4). The BV concentration (11 μM) required for half-maximal fluorescence in cells expressing iRFP was ~ 7 -fold lower than the concentration used in my screen for permuted variants (Figure 4.5). The majority of the permuted iRFP (85%) required a similar (or lower) amount of BV for half-maximal fluorescence, 4 to 15 μM (Figure 4.5). The remaining variants, which all arose from backbone fission in the GAF domain that binds BV, required between 18 and 25 μM BV for half-maximal fluorescence. These findings suggest that the large variability in whole cell fluorescence does not arise because circularly permuted iRFP differ in their BV-binding affinity.

4.3.6 Relationship between translation initiation and fluorescence

Previous studies have shown that genetic context can alter the strength of translation initiation from a single bacterial RBS by more than two orders of magnitude (Salis et al., 2009; Lou et al., 2012; Mutalik et al., 2013), and vectors in my library express permuted iRFP with sequences that vary adjacent to the same RBS. To examine whether permuted

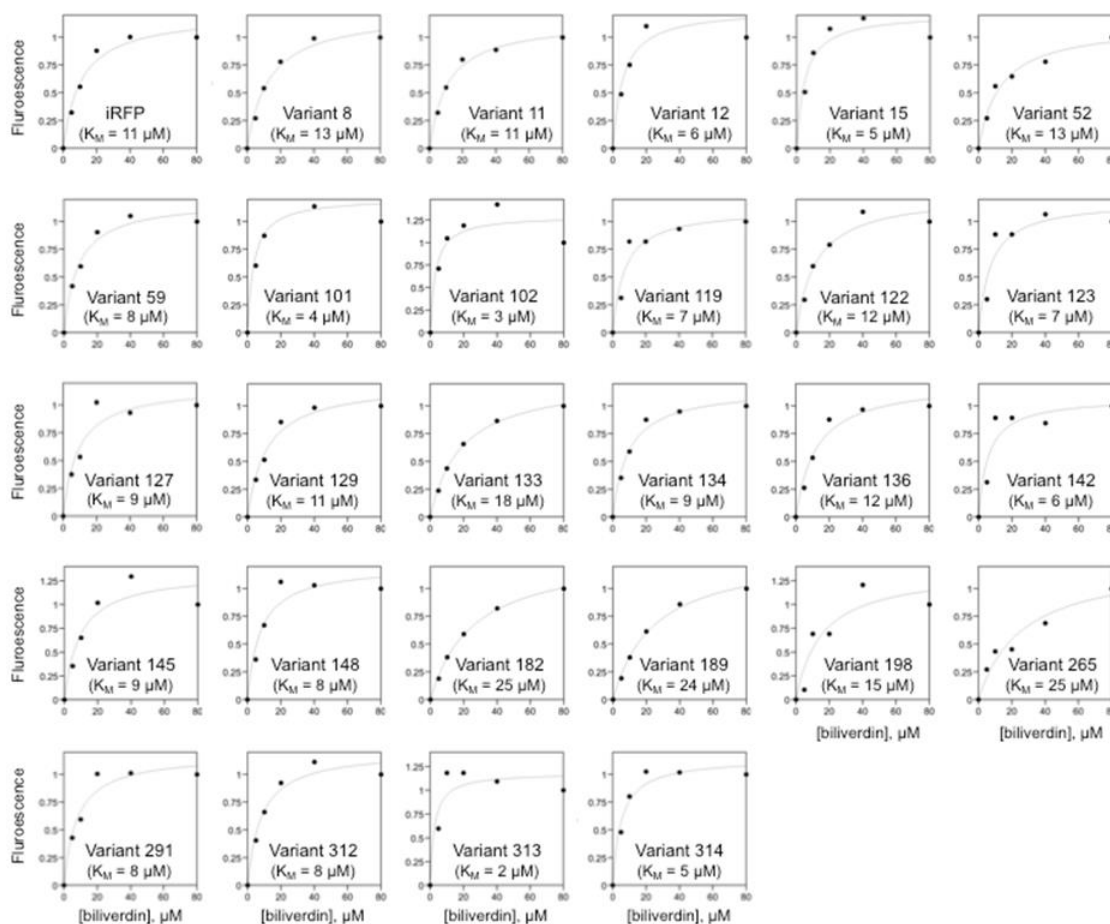


Figure 4.4 Effect of BV concentration on *E. coli* fluorescence expressing circularly permuted iRFP. The fluorescence ($\lambda_{\text{ex}} = 690 \text{ nm}$; $\lambda_{\text{em}} = 715 \text{ nm}$) of *E. coli* XL1 Blue expressing different circularly permuted iRFP was measured in cells grown in medium supplemented with varying BV concentrations (0, 5, 10, 20, 40, and 80 μM) and 1 mM arabinose. Fluorescence that had been normalized to the maximum signal obtained with each variant was fit to a hyperbolic saturation function to obtain the apparent affinity (K_M) within whole cells.

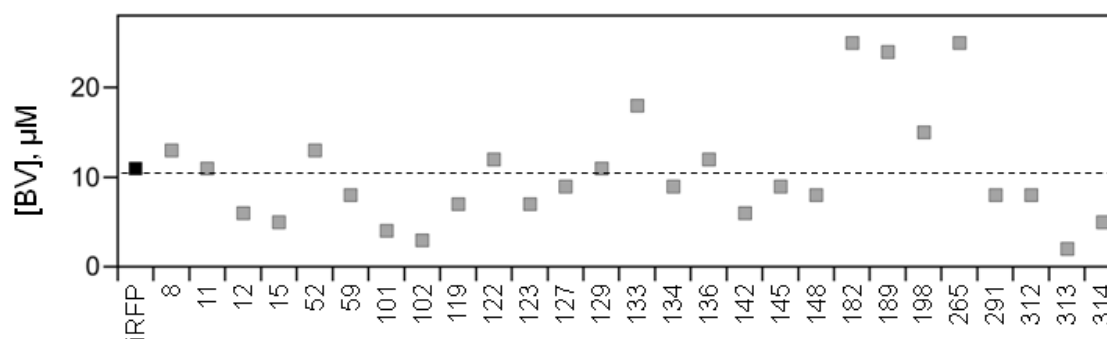


Figure 4.5 Trend of BV dependence among circularly permuted iRFP. Half maximal signal (K_M) for each variant was plotted against variants. Black filled square and grey filled squares represent iRFP and variants respectively. Variants below and above the dash line represent having improved and decreased BV affinity respectively.

gene sequences influence protein expression from the RBS in my permuteposon, I analyzed how permuted iRFP vary in their translation initiation rates in *E. coli* using a thermodynamic model for translation initiation (Salis et al., 2009). These calculations revealed that the calculated translation initiation rates for all possible circularly permuted iRFP vary by over three orders of magnitude when expressed using the RBS present in permuteposon P4 (Figure 4.6A). To test whether the fluorescent signals obtained with my original clones correlates with calculated translation initiation, I compared the translation initiation rates for each variant to the whole cell fluorescence signals. This analysis revealed that variants with the smallest relative translation initiation rates also display the weakest whole cell fluorescence when measured at 23°C and 37°C (Figure 4.6B). The dispersion of fluorescence at low calculated rates was also smaller than that at higher levels (*e.g.*, above calculated relative values of 1000). These findings suggest that the variability in the production rates of permuted iRFP contributes to the dispersion in fluorescence signals observed in whole cell measurements. The large dispersion of signals at higher calculated translation initiation rates suggests that protein folding and maturation vary across the proteins with similar translation initiation rates.

The close proximity between the RBS and the first codon in each circularly permuted gene within my permuteposon is expected to underlie the variability in translation initiation because each permuted gene presents a different genetic context to the same RBS. To minimize differences in protein synthesis that arise from this genetic context, I created vectors that express each of my circularly permuted iRFP with the same large N-terminal affinity tag (37 residues) that maintains the translation initiation site in a

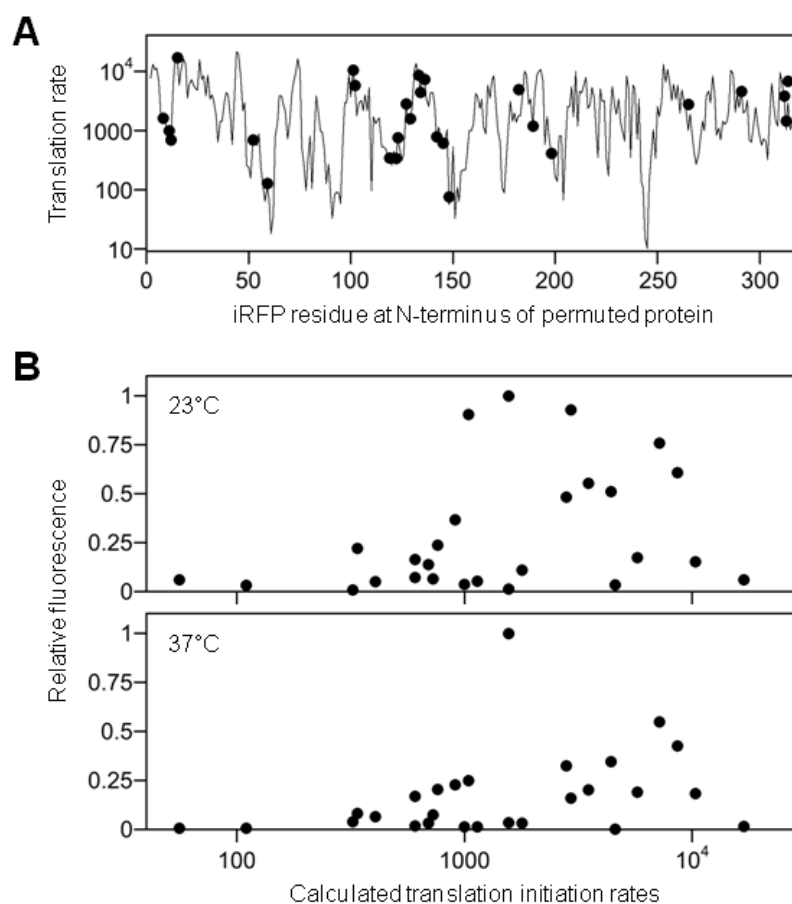


Figure 4.6 Thermodynamic analysis of permuted iRFP translation initiation. (A) Translation rates of all circularly permuted iRFP (line) are compared with the rates for fluorescent variants (●). (B) The relative whole cell fluorescent signals of circularly permuted iRFP at 23°C and 37°C are compared with calculated rates. Spearman's rank correlation coefficients (R_{SR}) at 23°C ($R_{SR} = 0.403$) and 37°C ($R_{SR} = 0.384$) yielded significant correlations (two-tailed t-test; $P < 0.05$).

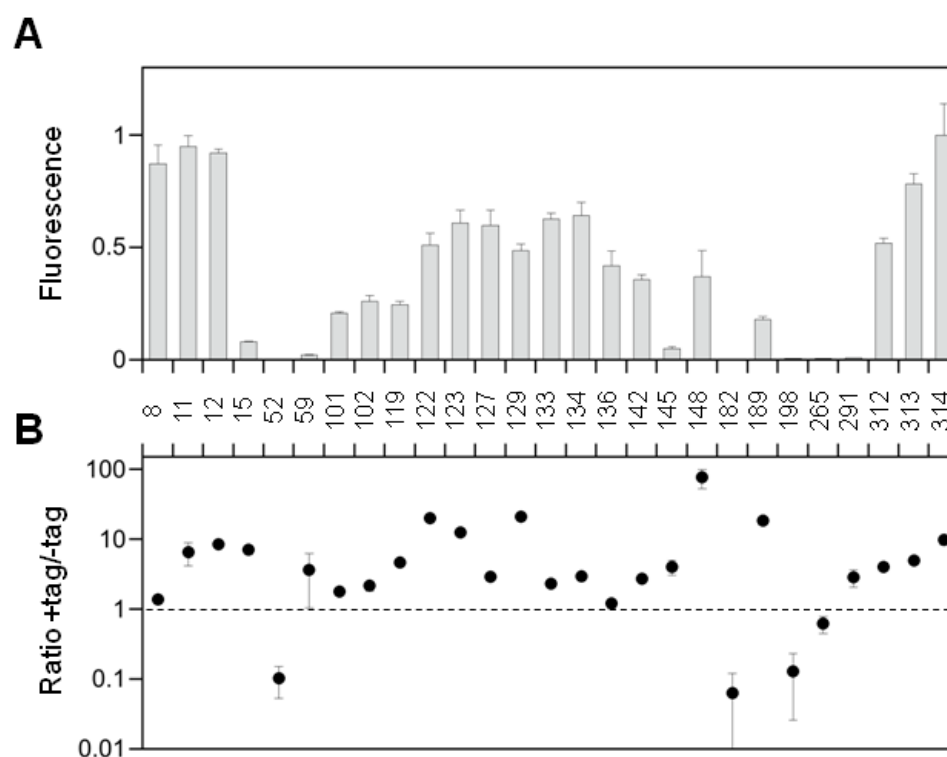


Figure 4.7 Addition of an N-terminal affinity tag enhances whole cell fluorescence. **(A)** The fluorescence ($\lambda_{\text{ex}} = 690 \text{ nm}$; $\lambda_{\text{em}} = 715 \text{ nm}$) of *E. coli* XL1 Blue expressing different circularly permuted iRFP with His tags was measured at 37°C and normalized to cell density. The signal from each variant is reported as the fraction of the signal obtained with the variant displaying the highest fluorescence. **(B)** The ratio of fluorescence measured for circularly permuted iRFP with a His tag (+tag) to same variant without a tag (-tag). The dashed line represents the ratio expected if expression does not vary with tag addition. Error bars represent $\pm 1\sigma$ from three or more experiments.

constant genetic context. When expressed from these vectors, permuted variants with new N- and C-termini proximal to the domain linker displayed the highest whole cell fluorescence (Figure 4.7). The fluorescence of each variant was also compared in the presence and absence of the affinity tag. A majority of the variants (85%) displayed an increased signal in the presence of the N-terminal tag, indicating that this expression strategy yielded higher steady-state levels of each fluorescent, BV-bound variant. To better understand the signal variability observed under these expression conditions, I analyzed the steady-state levels of a subset of these circularly permuted variants (Figure 4.8), focusing on the subset of circularly permuted iRFP that displayed enhanced fluorescence at 37°C compared with that observed at 23°C. Immunoblot analysis revealed that the steady-state levels of these circularly permuted varies up to 10-fold from that observed with iRFP. When whole cell fluorescence was normalized to protein expression in matched samples (Figure 4.9), a majority of the permuted variants displayed a signal that was similar to that measured with native iRFP. However, iRFP-198 displayed a signal that was 25-fold lower than iRFP, consistent with the low signal obtained in bacterial cells. In addition, iRFP variants 12 and 102 exhibited signals that were higher than iRFP, 1.94- and 4.38-fold respectively.

4.3.7 *In vitro* analysis of circularly permuted iRFP

My finding that permuted iRFP display a range of protein-normalized fluorescence signals in whole cell measurements suggested that permutation affects the fraction of iRFP that folds into a fluorescence-competent conformation and/or quantum yield. To directly address the effects of permutation on quantum yield, I purified two circularly

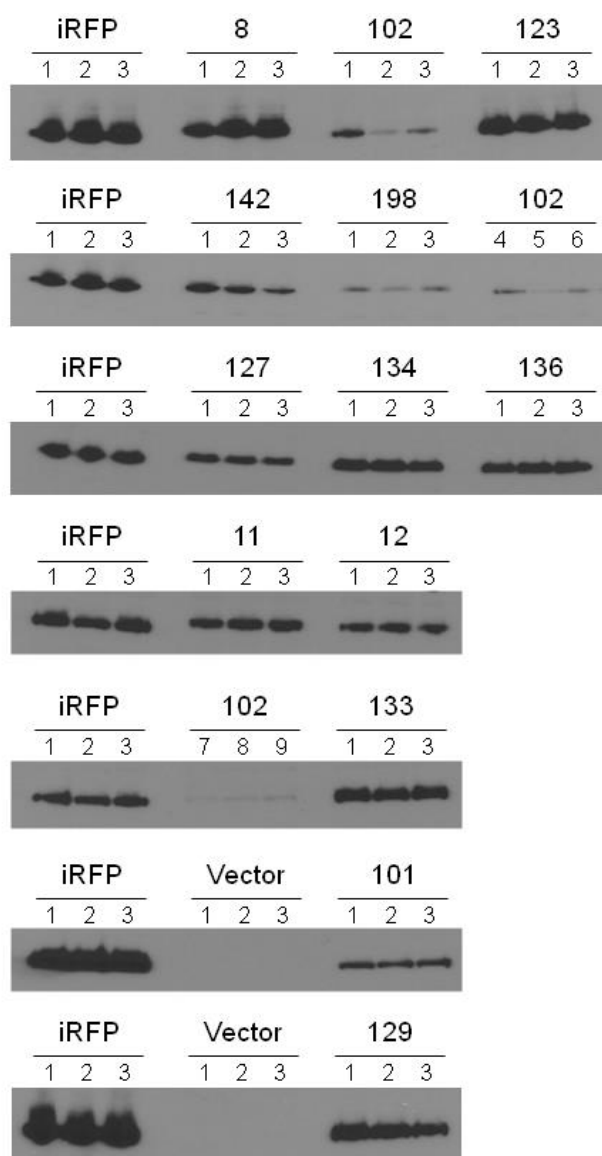


Figure 4.8 Immunoblot analysis of permuted protein expression. *E. coli* XL1-Blue transformed with pBAD vectors that express circularly permuted iRFP were subjected to Western immunoblot analysis using an anti-(His)₆ antibody. To establish the relative expression of each permuted variant, cells expressing unmodified iRFP was included as a control. Cells transformed with a pBAD vector lacking a BphP were included as a negative control. For each circularly permuted variant (8, 11, 12, 101, 102, 123, 127, 129, 133, 134, 136, 142, 198) experiments were performed using three or more samples derived from distinct colonies transformed with the same plasmid. All blots contained unmodified iRFP to allow for comparisons between different blots.

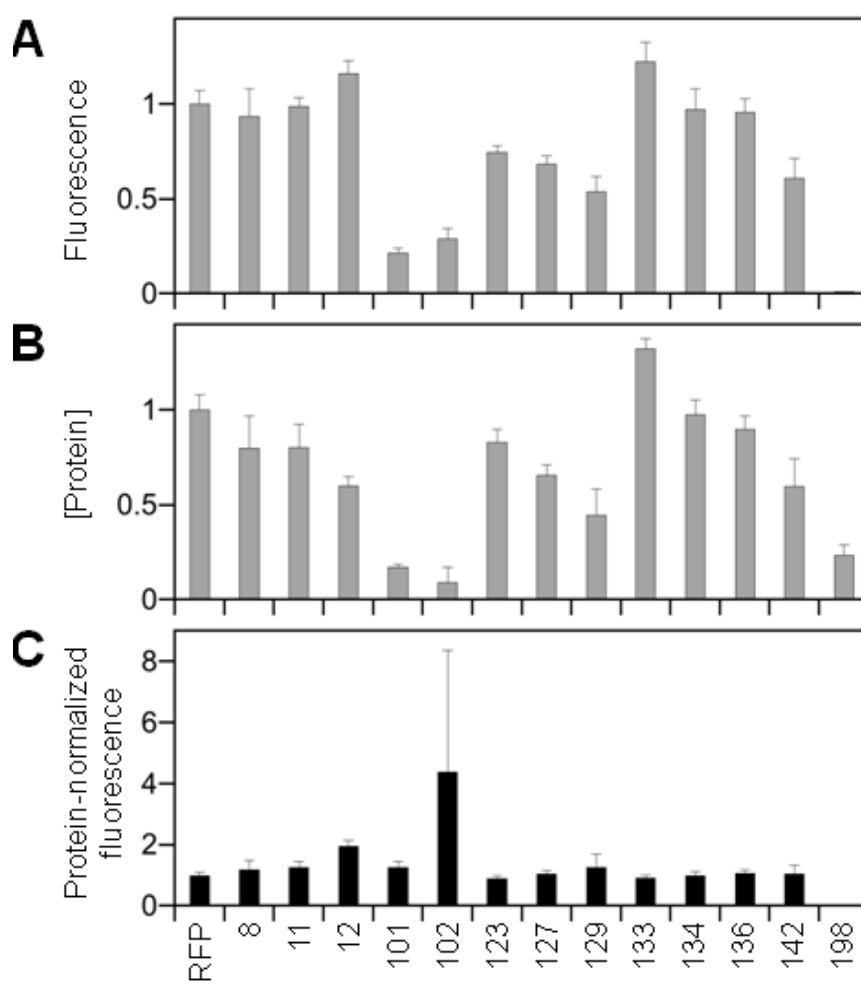


Figure 4.9 Protein-normalized fluorescence in *E. coli*. **(A)** Fluorescence of each circularly permuted iRFP upon expression is shown relative to that of cells expressing wild-type iRFP. Whole cell fluorescence was measured at 37°C and normalized to cell density. **(B)** The relative expression of the circularly permuted iRFP variants of panel A. ImageJ was used to quantify the levels of the variants in each blot, which are reported relative to the level of the wild-type iRFP. **(C)** The ratio of fluorescence to the level of each protein was calculated using data acquired from matched samples. Error bars represent $\pm 1\sigma$ calculated using three or more independent measurements.

permuted iRFP that displayed improved signals (variants 12 and 102), two with iRFP-like signals (variants 129 and 133) and iRFP. Proteins were purified using a combination of affinity and reverse phase chromatography. As observed with iRFP previously, I found that the reverse phase column yielded two populations of proteins, one that eluted as a monodisperse peak at low ammonium sulfate (~5%) and one that eluted when the column was washed with buffer lacking ammonium sulfate. All of my biochemical characterization was performed using the recombinant proteins that eluted in the presence of ammonium sulfate as monodisperse peaks.

I first compared the spectral properties of each permuted protein with iRFP (Figure 4.10). Like iRFP, all four of the permuted variants displayed three absorbance peaks with maxima at similar wavelengths (280, 391, and 693 nm). In addition, all four permuted variants displayed maximal excitation and emission wavelengths that mirrored iRFP (Table 4.2). Furthermore, all four variants displayed quantum yields that were not significantly different from the value obtained with iRFP (Table 4.2). These findings suggest that permutation does not significantly alter the chromophore environment.

I next sought to determine how permutation affects protein structure, since backbone fission sites within the dimerization interface of *Deinococcus radiodurans* BphP could disrupt oligomerization. Variants 129 and 133 introduce new termini that map onto the dimer interface in the BphP crystal structure (Figure 4.11A), while variants 12 and 102 have termini that map to sites distal from the dimer interface. To test whether these proteins display altered structures, I analyzed their electrophoretic mobility using native PAGE (Figure 4.11B). This analysis revealed that three variants (102, 129 and 133) display increased mobility compared with iRFP. To test if this altered mobility arises

Protein	Absorbance Maximum (nm)	Excitation Maximum (nm)	Emission Maximum (nm)	Oligomeric State	Quantum Yield	Extinction Coefficient (M ⁻¹ cm ⁻¹)	Conformational stability in GndHCl, [D] _{50%} (M)
iRFP	693	693	716	dimer	0.18 ±0.02	98144±19953	2.85
12	693	694	715	dimer	0.19 ±0.03	96417±14038	2.95
102	693	693	716	dimer	0.20 ±0.03	103185±20266	2.57
129	694	694	716	dimer	0.18 ±0.03	109874±16286	2.79
133	692	692	714	dimer	0.18 ±0.03	96325±18172	2.74

Table 4.2 *In vitro* properties of circularly permuted iRFP. Purified proteins were used for measuring *in vitro* properties of the variants. Measured properties were absorbance maximum, excitation maximum, emission maximum, oligomeric state, quantum yield, extinction coefficient and conformational stability.

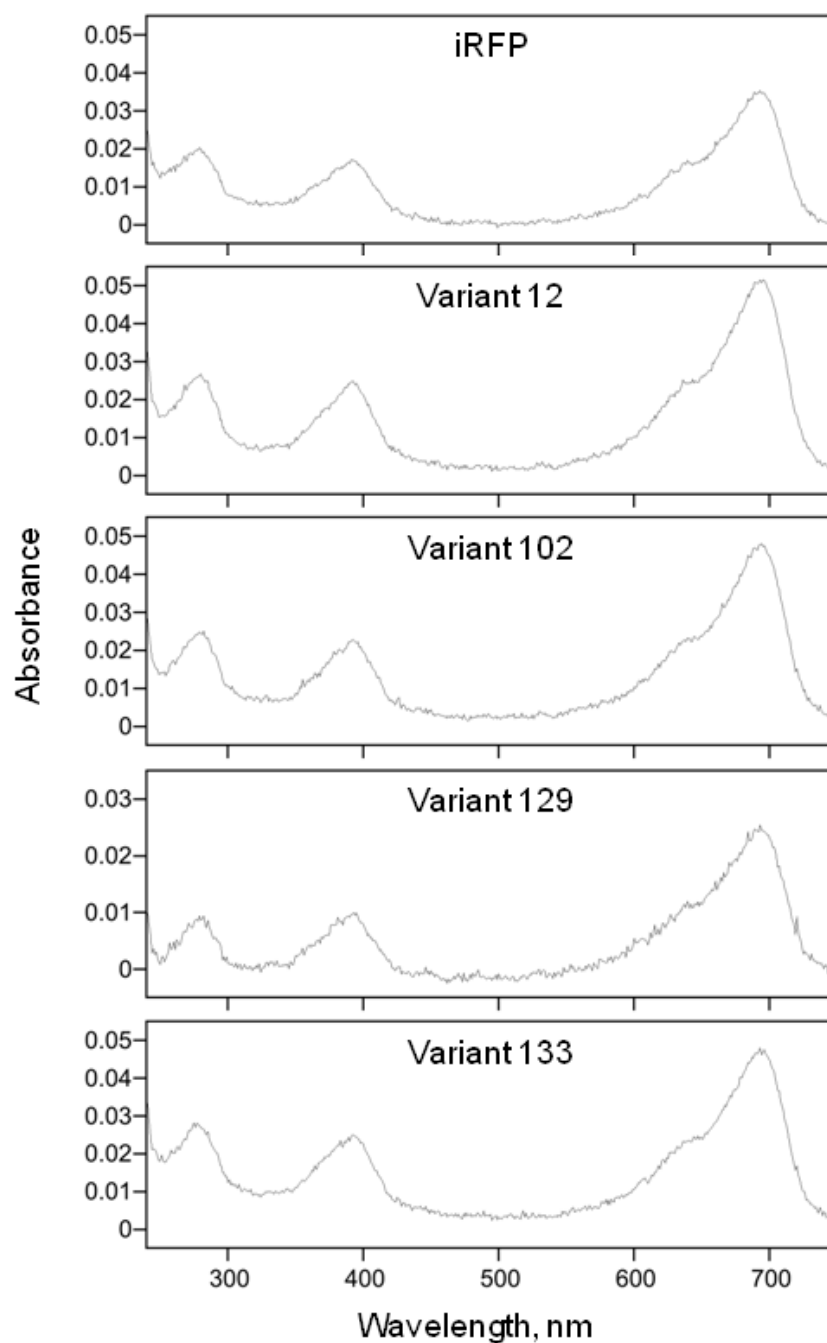


Figure 4.10 Absorbance of circularly permuted iRFP. Data acquired using purified proteins yielded similar absorbance peaks as native iRFP. A comparison of the ratio of the absorbance arising from BV (391 nm) and total protein (280 nm) yielded ratios for each permuted variant (0.9-1.1), which were similar to unmodified iRFP (0.85).

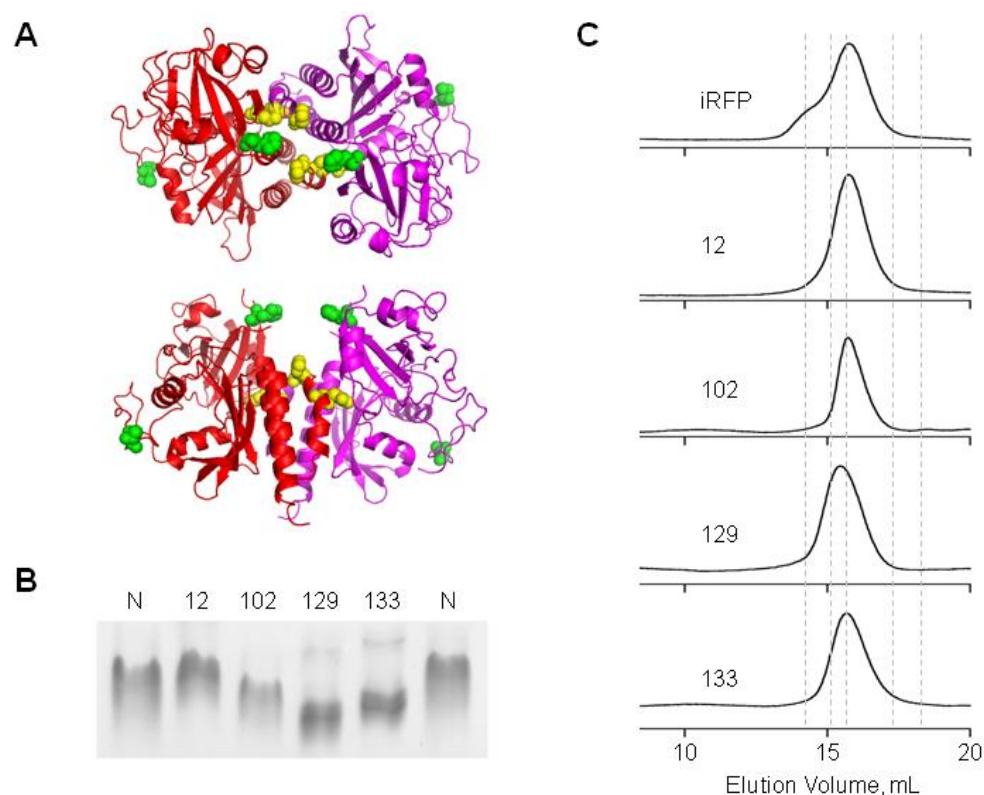


Figure 4.11 Effect of permutation on iRFP structure. (A) Dimer interface diagrams showing interaction of Bphp monomers (red and purple). Green represents variant 12 and 102 whereas yellow represents variant 129 and 133. (B) Native-PAGE of purified circularly permuted iRFP variants (12, 102, 129 and 133) and wild-type iRFP (N). (C) Gel filtration analysis of selected purified circularly permuted iRFP using Superdex 200 10/300 GL column chromatography using PBS buffer at 4°C. The dashed lines represent the volumes where the standards beta amylase (200 kDa), alcohol dehydrogenase (150 kDa), bovine serum albumin (66 kDa), carbonic anhydrase (29 kDa), cytochrome C (12.4 kDa) eluted. Number represents the first residue of the new circularly permuted iRFP.

because oligomerization is disrupted in these purified variants, I used an analytical gel filtration column to compare the oligomeric state of iRFP with variants 12, 102, 129, and 133 (Figure 4.11C). As previously observed, iRFP elutes as a monodisperse peak with an apparent MW (72,443 Da) that is 1.90 fold greater than the calculated MW of the iRFP polypeptide. In addition, all of the recombinant permuted proteins elute at volumes consistent with oligomerization. However, one variant (129) displayed an apparent molecular weight that was greater than iRFP (91,201 Da). This permuted iRFP also displayed the largest change in electrophoretic mobility with native PAGE. Taken together, these findings suggest that circular permutation affects iRFP shape but not oligomerization (Table 4.2).

Previous studies have shown that permutation can increase or decrease protein stability (Hultqvist et al., 2012). To examine how permutation affects iRFP stability, I compared the equilibrium unfolding of each purified permuted proteins with iRFP. Unfolding was monitored by measuring near infrared fluorescence signal as a function of the chemical denaturant guanidiniumhydrochloride (GndHCl). GndHCl-induced equilibrium unfolding of iRFP displayed a single cooperative transition with a midpoint at 2.85 M (Figure 4.12). The concentration of guanidine required for half-maximal unfolding of variants 102, 129 and 133 was lower (2.57, 2.79, and 2.74 M, respectively), indicating that these variants are destabilized compared with iRFP (Figure 4.12 and Table 4.2). In contrast, a higher concentration (2.95 M) of GndHCl was needed to achieve half maximal unfolding of variant 12 (Figure 4.12).

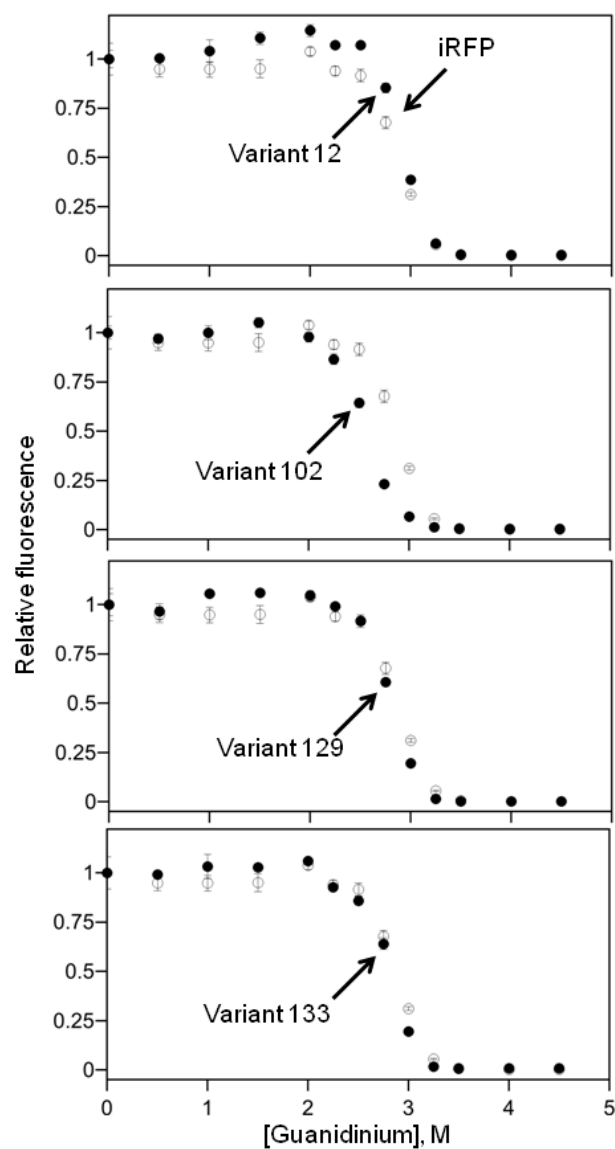


Figure 4.12 Equilibrium unfolding of circularly permuted iRFP. The effect of GndHCl concentrations on circularly permuted iRFP fluorescence (●) is compared with unmodified iRFP (○). For variant 12, GndHCl concentrations >1 M and <3.25 displayed signals that were significantly higher than iRFP (two-tailed t-test; $P < 0.05$). Error bars represent $\pm 1\sigma$ calculated using three or more measurements.

4.3.8 Fluorescence in mammalian cells

Although my initial analysis identified permuted iRFP with altered contact order that fold and fluoresce in bacteria, it was unclear if these variants would fold in a mammalian environment where the rate of translation is different. In addition, it wasn't clear if my variants would retain iRFP-like affinity for BV, which is sufficient to fluoresce in tissue culture cells without addition of exogenous BV (Filonov et al., 2011). To test these questions, I created mammalian constructs for constitutively expressing seven of my permuted iRFP, transiently transfected these plasmids into HeLa cells with a green fluorescent protein transfection control, and examined whole cell green and near infrared fluorescence using flow cytometry (Figure 4.13). A comparison of GFP-normalized near infrared fluorescence revealed that one of my permuted iRFP display enhanced whole cell fluorescence compared with iRFP (Figure 4.13), one variant displayed similar fluorescence as iRFP, four variants displayed ~25% of iRFP fluorescence, and one variant lacked detectable near infrared fluorescence. While bacterial measurements predicted that one variant should have low fluorescence (variant 198), the signals obtained with the other variants did not correlate across bacterial and mammalian cells. The brightest variants in tissue culture were variants 11 and 12, while the brightest variant in bacteria at 37°C was variant 133.

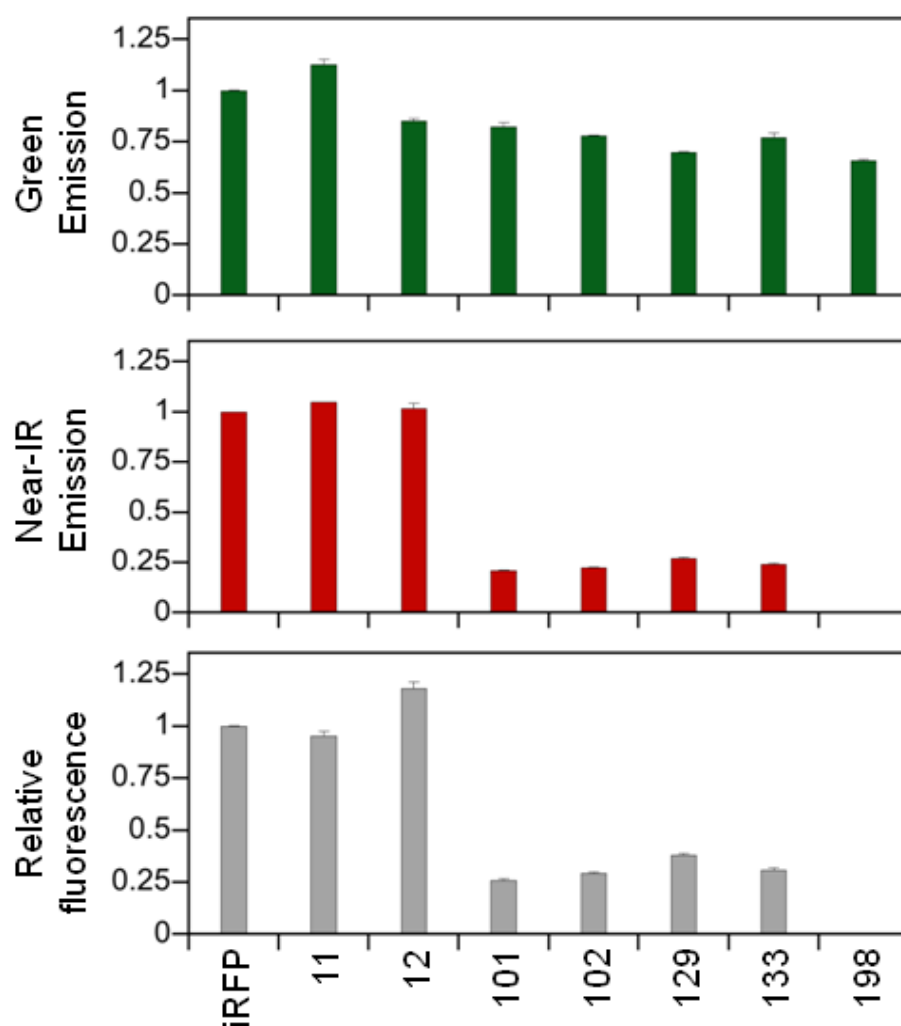


Figure 4.13 Flow cytometry analysis of tissue culture cells without exogenous BV. Green and near infrared fluorescence emission of HeLa cells transiently transfected with a pair of vectors that express GFP and select circularly permuted iRFP variants. For each variant near infrared fluorescence signal was normalized to the GFP transfection control to obtain relative fluorescence. Cells were grown in the medium that had not been supplemented with BV. Error bars represent $\pm 1\sigma$ calculated from two independent measurements. Variant 12 displayed a GFP-normalized signal that was significantly higher than iRFP (two tailed t test; $p < 0.05$).

4.4 Discussion

My results provide direct evidence that knotted BphP are tolerant to a wide range of sequence rearrangements that change contact order. High-throughput screening revealed that 8% of all possible circularly permuted iRFP display detectable near infrared fluorescence. This number is thought to represent a lower bound on the fraction of permuted iRFP that fluoresce, since my screening was only predicted to sample 58% of the possible permuted iRFP. Because retention of fluorescence requires a bound cofactor within a similar structural environment as iRFP, my findings suggest that many of these permuted proteins display iRFP-like structures. However, it is not clear which of these permuted variants acquire a knot like iRFP. My *in vitro* biochemical studies revealed parent-like stability and spectral properties, suggesting that the environment of the BV chromophore is not disrupted by permutation, while native gel analysis suggested that some variants may have subtle structural differences from iRFP. Within the BphP family there is evidence that some family members lack protein knots. Sequence comparisons have shown that the cyanobacteriochromes *AnPixJ* and *TePixJ* lack the loop that serves as the knot lasso (Narikawa et al., 2008; Ishizuka et al., 2007). In addition, a recent study revealed that the BphP knot can be removed through truncation without disrupting cofactor binding (Rumyantsev et al., 2015). However, unlike my circularly permuted iRFP, which display parent-like excitation and emission wavelengths, truncated BphP become blue shifted. These differences suggest that my permuted variants may still contain a knot. However, structural studies will be required to determine how the topology of my permuted proteins relate to iRFP, and if variants arising from backbone

fission in the domain linker are able to thread the full PAS domain through the loop to mature into a folded and fluorescent protein.

My whole cell fluorescence measurements revealed that all of my permuted variants display excitation and emission maxima that are similar to iRFP. However, at each temperature, I observed a large variation in fluorescence intensities even though each variant was transcribed and translated using the same promoter and RBS. In past studies, circular permutation has been shown to alter protein folding (Lindberg et al., 2002; Nobrega et al., 2014), ligand binding (Cheltsov et al., 2001; Qian et al., 2005), quantum yield (Topell et al., 1999) and expression (Jones et al., 2016). My results suggest that differences in permuted iRFP signal in *E. coli* do not arise from changes in BV binding affinity. The BV concentration required for half-maximal fluorescence was uniformly lower than the concentration used in my assays. However, expression did vary across each permuted variant. Thermodynamic calculations analyzing the effects of different permuted iRFP genes on the RBS in my permuteposon revealed a correlation between fluorescence and calculated translation initiation rates. In addition, whole cell fluorescence signals normalized to permuted protein accumulation revealed signals that were more consistent with iRFP. This observation emphasizes the importance of considering steady-state protein levels when using whole cell assays to analyze the effects of permutation on a protein.

Among the fluorescent variants discovered in my screen, variant 12 displayed significantly higher protein-normalized fluorescence compared with iRFP. Biochemical analysis of variant 12 did not reveal any differences in spectral properties from iRFP. This finding implicates improved maturation as the underlying cause of the enhanced

fluorescence. In my biochemical studies, I obtained evidence that only a fraction of the iRFP translated in *E. coli* matures into a fluorescent protein. When I subjected recombinant iRFP to reverse phase chromatography, I found that it eluted at two different salt concentrations and noted that these differed in their fluorescent signals as previously observed (Auldridge et al., 2012). Variant 12 also eluted at two different salt concentrations suggesting that only a subset of translated variant 12 matures into a fluorescent protein. However, my whole cell fluorescence measurements suggest that a greater fraction of this variant fully matures. Mapping variant 12 onto the *D. radiodurans* BphP structure (Wagner et al., 2005) suggests a mechanism for enhanced maturation. Because this permuted protein arises from fission within the N-terminal tail, near the BV attachment site at residue 15, it has fewer residues to thread through the loop that serves as the iRFP knot lasso. In future studies, the effects of permutation on protein maturation could be evaluated by analyzing protein folding and fluorescence within a cell free transcription and translation system. A recent study demonstrated that this approach can be used to study cofactor binding to a protein using the heme-binding myoglobin (Samuel et al., 2015). Using myoglobin orthologs of varying stability, this study observed a correlation between the levels of holo-myoglobin and the protein stability in the absence of cofactor (Samuel et al., 2015). This trend is consistent with my finding that variant 12 displays the highest stability among the permuted proteins purified, although my analysis was performed with BV-bound variants.

Although the permuted iRFP discovered in my screen did not yield variants with dramatic improvements in stability and quantum yield, my results have implications for protein design. Circularly permuted proteins are increasingly being used to design protein

switches through domain insertion, where an ensemble of permuted protein variants are randomly inserted at different locations within the backbone of a second protein to create new proteins (Ostermeier et al., 2009). Domain insertion studies have shown that a subset of the variants created in such libraries display switching behavior, where ligand binding within one domain modulates the structure and function of the second domain (Baird et al., 1999; Guntas et al., 2005). In fact, high signal-to-noise switches that couple the ligand-binding to the fluorescence of different GFP family members have been created by inserting circularly permuted fluorescent proteins into periplasmic binding proteins (Marvin et al., 2011). The circularly permuted iRFP discovered in my screens should be useful for creating similar biosensors, albeit with emission in the near infrared region which is better suited for tissue imaging (Filonov et al., 2011). All of my iRFP display parent-like BV affinity, which is sufficient to fluoresce in tissue culture without the addition of biliverdin. This suggests that biosensors created using these variants may have sufficient BV affinity to fluoresce in mammalian cells without requiring the addition of exogenous BV. The extent to which my permuted variants will fold, bind BV, and fluoresce upon insertion into other proteins will require further studies. If formation of a knot is required for permuted iRFP fluorescence, then it is possible that some of these iRFP variants could experience folding challenges upon domain insertion.

By using a circular iRFP gene for library construction, I was able to create vectors that express different permuted variants of iRFP using a single MuA reaction. This modified approach is quicker than the original protocol described for PERMUTE because it avoids manipulations required after the MuA reaction (Mehta et al., 2012), which take another day to perform. This modified PERMUTE protocol will simplify future studies

that seek to create vectors libraries that express different circularly permuted variants of other proteins. My results show that this simplified protocol can be used with permuteposons that contain hybrid transposase and ribosomal binding sites that only amend two extra amino acids to protein termini (Jones et al., 2016). This protocol should also be compatible with other permuteposons that have been developed for PERMUTE, including those that amend peptides of varying size to the N-termini of permuted proteins (Mehta et al., 2012). Beyond simplifying the construction of permuted protein libraries, the approach described herein has the advantage that it creates permuted proteins with well-defined sequence diversity and avoids random deletions that arise with previously described methods (Graf et al., 1996; Hennecke et al., 1999). Whether or not this approach comprehensively samples all possible permuted protein variants will require future studies that use deep sequencing to analyze library diversity, similar to that described for libraries created using random mutation (Firnberg et al., 2012).

Chapter 5

Conclusions and future directions

5.1 Knotted IFP are sensitive to backbone fission

Knotted proteins are now widely accepted as naturally occurring proteins. However, the structural and function role(s) for knot are still not fully appreciated. The work described herein shows that knotted proteins are quite robust to mutational lesions that cleave the backbone and alter contact order. In my study of random IFP fragmentation, I found that this protein displays a low tolerance to fragmentation. However, addition of the associating peptides to the termini of each fragment led to the discovery of 13 fragmented IFP. Removing the associating peptides from these 13 split proteins uniformly diminished the fluorescence of the fragmented IFP. This could be due to the fact that unassisted IFP fragment cannot form the proper knot, or because the fragments simply do not associate strongly. Whether these split IFP still form a knot is not known. My split protein results can be contrasted with previous studies in the Silberg lab. Analysis of the tolerance of an adenylate kinase to fission revealed that a thermophilic protein retains function 40% of the time, while a mesophilic ortholog retained function ~6% of the time (Segall-Shapiro et al., 2011). The frequency of fluorescent two fragment IFP described in this thesis (4%) is similar to this latter value, although assistance was required from associating proteins.

5.2 Split IFP should be useful for studying protein-protein interactions *in situ*

The dependence of my fragmented IFP on associating proteins for maximal

fluorescence suggests that these split proteins will be useful for reporting on other protein-protein interactions like previously described bimolecular fluorescence complementation assays (Hu et al., 2002; Wilson et al., 2004; Kerppola et al., 2009). The longer wavelengths of light used to excite IFP penetrates tissues to a greater extent than the wavelengths used to excite GFP-family members, suggesting that these split IFP should be useful as reporters of molecular interactions within animal models like IFP (Shu et al., 2009). These split proteins should also be useful in anaerobic environments where GFP-family members cannot be used. Whether or not they are reversible will require further investigation. In the case of split GFP, fragment association is irreversible which hinder its use in aggregations studies. The fragmented IFP identified by laboratory evolution will also be useful for studies examining the spatial and temporal distribution of promoter activities within different cells within animals. Such measurements could be used to analyze the cells in complex populations where pairs of promoters are active. Such information can be coupled to emerging chemical biology tools, such as two-input AND gates that control metabolic labeling of newly synthesized proteins (Mahdavi et al., 2013). Such methods enable proteomic measurements with improved temporal and spatial resolution (Yuet et al., 2014). IFP fragment complementation can be used to report on the subset of cells within an organism (bacterial or eukaryotic) where a given pair of promoters is active, information that cannot be obtained using pairs of near-infrared fluorescent protein reporters because their excitation and emission spectra overlap (Shcherbakova et al., 2013). However, IFP has a low quantum yield, and this property could limit the application of the split IFP under conditions where fragments are expressed at low levels. Whether or not simultaneous expression of both IFP fragments is

required to achieve optimal fluorescence will require future studies that examine how each fragment degrades in the absence of other fragment.

5.3 A generalizable method for discovering split protein that depend on protein-protein interactions

My results show that using laboratory evolution to discover split protein has fundamental challenges associated with protein expression. In bacteria, translation initiation is determined by the context of an RBS. With split protein, different gene fragments adjacent to the same RBS change its context. To keep similar translation initiation rate across multiple variants in a vector library, it is best to have a similar sequence (~35 bp) next to RBS. In my split protein library approach, I found that protein expression can be made consistent by using a two step procedure. My results suggest that one should first screen for fragmented proteins that retain function upon fusion to a pair of interacting proteins like IAAL-E3 and IAAL-K3. Upon discovery, one should rescreen for loss-of-function upon removal of only one of the fused proteins, i.e., IAAL-E3, because it is located at the end of the first fragment and removal will not affect the context of any RBS. In contrast, removing both IAAL-E3 and IAAL-K3 could lead to false negatives due to changes in translation initiation rates. Future applications of this transposon mutagenesis approach are expected to simplify the discovery of fragmented protein devices that couple the function of other proteins to protein-protein interactions (Michnick et al., 2007). However, deep sequencing may be required for analyzing libraries that contain and lack fusion proteins, since some fission sites may be non-disruptive to protein function in the absence of protein fusions. To identify those that

require fusion to proteins that interact, it may be important to deep sequence libraries \pm fusions before screening and after screening to find the backbone locations that are most enriched \pm fusions. In cases where fused proteins have an interaction that is conditional, *e.g.*, dependent upon ligand (Remy et al., 1999) or cofactor (Hoff et al., 2009) binding, fragmented proteins will display three-input AND gate logic. The first two inputs will be the promoters controlling protein fragment expression, while a third input will be the post-translational event that promotes the protein-protein interaction.

5.4 Split proteins will be useful for studying DNA minivector delivery *in situ*

The protein fission strategy described here is expected to be useful for synthetic biology applications that involve DNA minivectors (and minicircles), small vectors that display amazing stability. These small DNA are being developed for gene therapy because of their ability to enter difficult-to-transfect mammalian cells and survive shear forces often associated with DNA delivery (Zhao et al., 2011; Catanese et al., 2012; Stenler et al., 2014). The beneficial properties of these small DNA are inversely correlated with DNA length, and only a limited number of natural proteins can be encoded by a single minivector at the sizes (<1,000 base pairs) where these small DNA display the greatest stabilities. The benefits of minivectors could be leveraged with larger proteins if those proteins were broken into smaller fragments that fit into minivectors having sizes <1,000 base pairs, because pairs or trios of minivectors encoding these fragments could be delivered to difficult-to-transfect mammalian cells to reconstitute a functional protein through fragment complementation (Figure 5.1). Because my results show that IFP can be fragmented into pieces that associate through the fusion of small

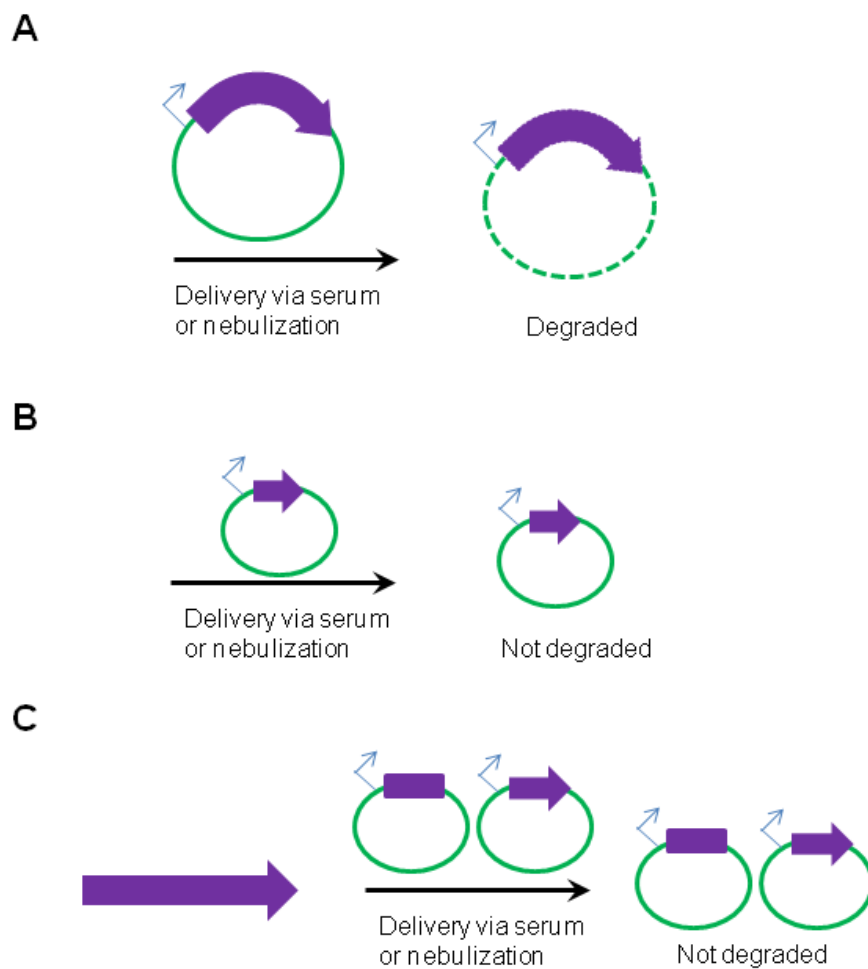


Figure 5.1 Split IFP for gene delivery using DNA minivectors. **(A)** Typical viral or plasmid vectors with gene of interest (purple) for gene therapy are larger than 5 kb, which gets degraded during delivery by shear force developed in blood or nebulization. **(B)** Alternatively, supercoiled minivectors are smaller vectors (>1200 bp) which survive shear forces but only deliver small genes. **(C)** To deliver larger genes (including IFP), split genes system can be utilized by cloning fragments into multiple minivectors for delivery. Using minivectors for delivering of split IFP one can test the dosage response of the split protein delivery for hard to transfect tissues and in deep tissues.

coils, I expect that the approach described herein will be useful for breaking natural proteins into pieces that can be encoded by minivectors, *e.g.*, delivery protein therapeutics in pieces.

5.5 Circularly permuted iRFP will be useful for domain insertion

Near infrared fluorescent proteins provide powerful imaging tools to decipher cellular activity in tissue of live animals. The variants discovered in this thesis work are poised to be used to create biosensors with a range of sensing functions. For example, the circularly permuted iRFP discovered in my thesis work will be useful for domain insertion. With domain insertion, a protein of interest is randomly inserted into the target protein, and screening (or selection) is used to identify functional proteins that display switching activity (Baird et al., 1999; Guntas et al., 2005). In some cases, domain insertion yields a pair of domains that fold independently and occasionally the function of one domain is regulated by the activity of the other protein. In domain insertion, the N- and C-termini of the inserted protein must be nearby one another or they will disrupt the folding of the other protein. This design constraint has limited previous applications of IFP/iRFP in domain insertion, since these proteins have N- and C-termini that are distant ($>20\text{\AA}$) from one another. My circularly permuted proteins should be compatible with domain insertion since permutation links the original distant N- and C-termini with flexible linker and creates new proximal N- and C-termini by backbone fission elsewhere in the protein (Figure 1.3 and Figure 5.2). In this study, I have generated 27 variants of circularly permuted variants, all of which should be useful for domain insertion. These can be inserted into any protein imaginable to couple IFP folding, cofactor binding and/or



Figure 5.2 Domain insertion using permuted iRFP. Domain insertion required closer N- and C-termini of the original protein. For those protein (including iRFP) that do not fit this criteria, circular permutation is performed to create closer N- and C-termini. In this process original N- and C-termini are connected through flexible peptide linker (blue line) and new N- and C-termini are created somewhere else in the protein. New N- and C-termini are connected to the other protein or domain (green) to create a domain inserted protein fusion which are useful for designing allosteric switches or sensors.

fluorescence to other proteins. For example, circularly permuted could be randomly inserted into proteins that bind specific metabolites, and these libraries could be screened for variants that display metabolite-dependent near infrared fluorescence.

5.6 Conclusions and Outlook

In this thesis, I provide the first evidence that a knotted protein can be split and permuted without disrupting structure and function using laboratory evolution. I have identified multiple split and permutation sites (13 and 27, respectively). Some of these backbone fission sites cluster near the interdomain linker between PAS and GAF domain, although some differences are observed in the sites discovered from backbone fission and permutation. Many questions remain about knotted BphP. It is not clear if my split IFP and permuted iRFP fold into a knot like full-length IFP and iRFP. Biophysical analysis will be required to answer this question. X-ray crystallography could be potentially used to directly probe the presence of the knot in the structure. However, it might be challenging to obtain crystals for these proteins, especially the split proteins which could be challenging to purify. NMR may be useful for structural analysis as well.

The stability of split IFP and permuted iRFP might be affected by backbone fission to differing extents. IFP is less stable than iRFP (Filonov et al., 2011), and this could explain why no two fragment IFP were discovered that displayed fluorescence in the absence of protein fusions. A laboratory evolution experiment subjecting iRFP to random fission might discover split variants that do not display the same dependence upon protein-protein interactions. For permuted variants of the iRFP, GndHCl unfolding of the purified variants did not show major difference in stability among the variants (Table

4.2), although one variant displayed a minor increase in stability. Whether the same is true when IFP is circularly permuted is not known. Additional studies will be required to determine if IFP or other BphP can be permuted at the same locations as iRFP without disrupting folding, structure and function. It is tempting to speculate that BphP, which sense light and use light dependent isomerization of their cofactor for signaling, could be permuted like iRFP and used in domain insertion experiments to create novel light switchable proteins.

References

- Auldridge, M. E., Satyshur, K. A., Anstrom, D. M., and Forest, K. T. (2012). Structure-guided engineering enhances a phytochrome-based infrared fluorescent protein. *J. Biol. Chem.* 287, 7000-7009.
- Baird, G. S., Zacharias, D. A., and Tsien, R. Y. (1999). Circular permutation and receptor insertion within green fluorescent proteins. *Proc. Natl. Acad. Sci. U.S.A.* 96, 11241-11246.
- Barnard, E., McFerran, N.V., Trudgett, A., Nelson, J., and Timson, D.J. (2008). Detection and localisation of protein–protein interactions in *Saccharomyces cerevisiae* using a split-GFP method. *Fungal Genetics and Biology*, 45, 597-604.
- Barondeau, D.P., Putnam, C.D., Kassmann, C.J., Tainer, J.A., and Getzoff, E.D. (2003). Mechanism and energetics of green fluorescent protein chromophore synthesis revealed by trapped intermediate structures. *Proc. Natl. Acad. Sci. U.S.A.* 100, 12111-12116.
- Bevis, B.J., and Glick, B.S. (2002). Rapidly maturing variants of the *Discosoma* red fluorescent protein (DsRed). *Nat. Biotech.* 20, 83-87.
- Bloom, J.D., Silberg, J.J., Wilke, C.O., Drummond, D.A., Adami, C., Arnold, F.H. (2005). Thermodynamic prediction of protein neutrality. *Proc. Natl. Acad. Sci. U.S.A.* 102, 606-11.
- Burbaum, J. J. and Schimmel, P. (1991) Assembly of a class I tRNA synthetase from products of an artificially split gene. *Biochemistry* 30, 319-324.
- Cabantous, S., Terwilliger, T.C., and Waldo, G.S. (2005). Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. *Nat. Biotech.* 23, 102-107.
- Cabantous, S., Nguyen, H.B., Pedelacq, J.D., Koraïchi, F., Chaudhary, A., Ganguly, K., Lockard, M. A., Favre, G., Terwilliger, T. C., and Waldo, G. S. (2013). A new protein–protein interaction sensor based on tripartite split-GFP association. *Sci. Rep.* 3, 2854.
- Catanese, D. J., Fogg, J. M., Schrock, D. E., Gilbert, B. E., and Zechiedrich, L. (2012). Supercoiled minivector DNA resists shear forces associated with gene therapy delivery. *Gene Ther.* 19, 94-100.
- Chalfie, M., Tu, Y., Euskirchen, G., Ward, WW., and Prasher, D.C. (1994). Green fluorescent protein as a marker for gene expression. *Science* 263, 802-805.

- Chapleau, R.R., Blomberg, R., Ford, P.C., and Sagermann, M. (2008). Design of highly specific and noninvasive biosensor for real-time *in vivo* imaging of mercury (II) uptake. *Protein Science* 17, 614-622.
- Cheltsov, A.V., Barber, M.J., and Ferreira, G.C. (2001). Circular permutation of 5-aminolevulinate synthase. Mapping the polypeptide chain to its function. *J. Biol. Chem.* 276, 19141-19149.
- Choe, W., Chandrasegaran, S., and Ostermeier, M. (2005). Protein fragment complementation in M.HhaI DNA methyltransferase. *Biochem. Biophys. Res. Commun.* 334, 1233-1240.
- Chudakov, D.M., Matz, M.V., Lukyanov, S., and Lukyanov, K.A. (2010). Fluorescent proteins and their applications in imaging living cells and tissues. *Physiology Rev.* 90, 1103-1163.
- Daugherty, A.B., Govindarajan, S., and Lutz, S. (2013). Improved Biocatalysts from a Synthetic Circular Permutation Library of the Flavin-Dependent Oxidoreductase Old Yellow Enzyme. *J. Am. Chem. Soc.* 135, 14425–14432.
- Ducrest, A., Amacker, M., Linger, J., and Markus Nabholz. (2002). Detection of promoter activity by flow cytometric analysis of GFP reporter expression. *Nucleic Acids Research* 30, e65.
- Feilmeier, B.J., Iseminger, G., Schroeder, D., Webber, H., and Phillips, G.J. (2000). Green fluorescent protein functions as a reporter for protein localization in *Escherichia coli*. *J. Bacteriology* 182, 4068-4076.
- Fery-Forgues, S., and Lavabre, D. (1999). Are fluorescence quantum yields so tricky to measure? A demonstration using familiar stationery products. *J. Chem. Educ.* 76, 1260-1264.
- Fields, S. and Song, O. (1989). A novel genetic system to detect protein–protein interactions. *Nature* 340, 245-246.
- Filonov, G.S., Krumholz, A., Xia, J., Yao, J., Wang, L.V., and Verkhusha, V.V. (2012). Deep-tissue photoacoustic tomography of genetically encoded iRFP probe. *Angew. Chem. Int. Ed. Engl.* 51, 1448-1451.
- Filonov, G.S., Piatkevich, K.D., Ting, L., Zhang, J., Kim, K., and Verkhusha V.V. (2011). Bright and stable near-infrared fluorescent protein for *in vivo* imaging. *Nat. Biotech.* 29, 757-761.
- Filonov, G. S. and Verkhusha, V. V. (2013). A near-infrared BiFC reporter for *in vivo* imaging of protein–protein interactions. *Chem. Biol.* 20, 1078-1086.

- Firnberg, E., and Ostermeier, M. (2012). PFunkel: efficient, expansive, user-defined mutagenesis. *PLoS ONE* 7, e52031.
- Fradkov, A.F., Verkhusha, V.V., Staroverov, D.B., Bulina, M.E., Yanushevich, Y.G., Martynov, V.I., Lukyanov, S., and Lukyanov, K.A. (2002). Far-red fluorescent tag for protein labeling. *Biochemical J.* 368, 17-21.
- Galarneau, A., Primeau, M., Trudeau, L.-E., and Michnick, S. W. (2002) Beta-lactamase protein fragment complementation assays as *in vivo* and *in vitro* sensors of protein protein interactions. *Nat. Biotechnol.* 20, 619-622.
- Giraldez, T., Hughes, T.E., and Sigworth, F.J. (2005). Generation of Functional Fluorescent BK Channels by Random Insertion of GFP Variants. *J. Gen. Physiol.* 126, 429-438.
- Giraud, E., Zappa, S., Vuillet, L., Adriano, J, Hannibal, L., Fardoux, J., Berthomieu, C., Bouyer, P., Pignol, D., and Vermeglio, A. (2005). A New Type of Bacteriophytochrome Acts in Tandem with a Classical Bacteriophytochrome to Control the Antennae Synthesis in *Rhodospseudomonas palustris*. *J. Biol. Chem.* 280, 32389-32397.
- Graf, R., and Schachman, H.K. (1996). Random circular permutation of genes and expressed polypeptide chains: application of the method to the catalytic chains of aspartate transcarbamoylase. *Proc. Natl. Acad. Sci. U.S.A.* 93, 11591-11596.
- Grünberg, R. and Serrano, L. (2010). Strategies for protein synthetic biology. *Nucleic Acids Res.* 38, 2663-2675.
- Guntas, G., Mansell, T.J., Kim, J.R. and Ostermeier, M. (2005). Directed evolution of protein switches and their application to the creation of ligand-binding proteins. *Proc. Natl. Acad. Sci. U.S.A.* 102, 11224-11229.
- Haglund E., Lindberg, M.O., and Oliveberg, M. (2008). Changes of protein folding pathways by circular permutation. *J. Biol. Chem.* 283, 27904-27915.
- Hastings, J.W. (1996). Chemistries and colors of bioluminescent reactions: a review. *Gene* 173, 5-11.
- Heim R., and Tsien, R.Y. (1996). Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. *Current Biology* 6, 178-182.
- Heim, R., Prasher, D.C., and Tsien, R.Y. (1994). Wavelength mutations and posttranslational autooxidation of green fluorescent protein. *Proc. Natl. Acad. Sci. U.S.A.* 91, 12501-12504.

- Hennecke, J., Sebbel, P., and Glockshuber, R. (1999). Random circular permutation of DsbA reveals segments that are essential for protein folding and stability. *J. Mol. Biol.* 286, 1197-1215.
- Hida, K., Won, S.Y., Di Pasquale, G., Hanes, J., Chiorini, J.A., and Ostermeier, M. (2010). Sites in the AAV5 capsid tolerant to deletions and tandem duplications. *Arch. Biochem. Biophys.* 496, 1-8.
- Hock, A.K., Lee, P., Maddocks, O.D., Mason, S.M., Blyth, K., and Vousden, K.H. (2013). iRFP is a sensitive marker for cell number and tumor growth in high-throughput systems. *Cell Cycle* 13, 220-226.
- Hoff, K.G., Culler, S.J., Nguyen, P.Q., McGuire, R.M., Silberg, J.J., and Smolke, C.D. (2009). *In Vivo* Fluorescent Detection of Fe-S Clusters Coordinated by Human GRX2. *Chem. Biol.* 16, 1299-1308.
- Hu, C.D. and Kerppola, T. K. (2003). Simultaneous visualization of multiple protein interactions in living cells using multicolor fluorescence complementation analysis. *Nat. Biotechnol.* 21, 539-545.
- Hu, C.D., Chinenov, Y., and Kerppola, T. K. (2002). Visualization of interactions among bZIP and Rel family proteins in living cells using bimolecular fluorescence complementation. *Mol. Cell* 9, 789-798.
- Hultqvist, G., Punekar, A.S., Morrone, A., Chi, C.N., Engstrom, A., Selmer, M., Gianni, S., and Jemth, P. (2012). Tolerance of protein folding to a circular permutation in a PDZ domain. *PLoS One* 7, e50055.
- Ishizuka, T., Narikawa, R., Kohchi, T., Katayama, M., and Ikeuchi, M. (2007). Cyanobacteriochrome TePixJ of *Thermosynechococcus elongatus* harbors phycoviolobilin as a chromophore. *Plant Cell Physiol.* 48, 1385-1390.
- Iwakura, M., Nakamura, T., Yamane, C., and Maki, K. (2000). Systematic circular permutation of an entire protein reveals essential folding elements. *Nat. Struct. Mol. Biol.* 7, 580-585.
- Jamroz, M., Niemyska, W., Rawdon, E.J., Stasiak, A., Millett, K.C., Sulkowski, P., and Sulkowska, J.I. (2015). KnotProt: a database of proteins with knots and slipknots. *Nucleic Acids Res.* 43, D306-D314.
- Joanna I. Sułkowska^{a,b,1}, Piotr Sułkowski^{c,d}, P. Szymczak, and Marek Cieplak. (2008). Stabilizing effect of knots on proteins. *Proc. Natl. Acad. Sci. U.S.A.* 105, 19714-19719.
- Jobsis, F.F. (1977). Noninvasive, infrared monitoring of cerebral and myocardial oxygen sufficiency and circulatory parameters. *Science* 198, 1264-1267.

- Johnsson, N. and Varshavsky, A. (1994). Split ubiquitin as a sensor of protein interactions *in vivo*. *Proc. Natl. Acad. Sci. U.S.A.* 91, 10340-10344.
- Jones, A. M., Mehta, M. M., Thomas, E. E., Atkinson, J. T., Segall-Shapiro, T. H., Liu, S., and Silberg, J. J. (2016). The structure of a thermophilic kinase shapes fitness upon random circular permutation. *ACS Synth. Biol.* (In press)
- Karimova, G., Pidoux, J., Ullmann, A., and Ladant, D. (1998). A bacterial two-hybrid system based on a reconstituted signal transduction pathway. *Proc. Natl. Acad. Sci. U.S.A.* 95, 5752-5756.
- Kerppola, T. K. (2009). Visualization of molecular interactions using bimolecular fluorescence complementation analysis: characteristics of protein fragment complementation. *Chem. Soc. Rev.* 38, 2876-2886.
- King, N.P., Jacobitz, A.W., Sawaya, M.R., Goldschmidt, L., and Yeates, T.O. (2010). Structure and folding of a designed knotted protein. *Proc. Natl. Acad. Sci. U.S.A.* 107, 20732-20737.
- Kneen, M., Farinas, J., Li, Y., and Verkman, A.S. (1998). Green fluorescent protein as a noninvasive intracellular pH indicator. *Biophys. J.* 74, 1591-1599.
- Lamparter, T., Michael, N., Mittmann, F., and Esteban, B. (2002). Phytochrome from *Agrobacterium tumefaciens* has unusual spectral properties and reveals an N-terminal chromophore attachment site. *Proc. Natl. Acad. Sci. U.S.A.* 99, 11628-11633.
- Landro, J.A., and Schimmel, P. (1993). Non-covalent protein assembly. *Curr. Opin. Struct. Biol.* 3, 549-554.
- Liang, C., and Mislow, K. (1995). Topological Features of Protein Structures: Knots and Links. *J. Am. Chem. Soc.* 117, 4201-4213.
- Lindberg, M., Tangrot, J., and Oliveberg M. (2002). Complete change of the protein folding transition state upon circular permutation. *Nat. Struct. Biol.* 9, 818-822.
- Lindenburg, L., and Merkx, M. (2014). Engineering genetically encoded FRET sensors. *Sensors* 14, 11691-11713.
- Lindhout, D. A., Litowski, J. R., Mercier, P., Hodges, R. S., and Sykes, B. D. (2004). NMR solution structure of a highly stable de novo heterodimeric coiled-coil. *Biopolymers* 75, 367-375.
- Lippincott-Schwartz, J., Snapp, E., and Kenworthy, A. (2001). Studying protein dynamics in living cells. *Nat. Rev. Mol. Cell Biol.* 2, 444-456.

- Litowski, J. R. and Hodges, R. S. (2002). Designing heterodimeric two-stranded alpha-helical coiled-coils. Effects of hydrophobicity and alpha-helical propensity on protein folding, stability, and specificity. *J. Biol. Chem.* 277, 37272-37279.
- Llopis, J., McCaffery, J.M., Miyawaki, A., Farquhar, M.G., Tsien, R.Y. (1998). Measurement of cytosolic, mitochondrial, and Golgi pH in single living cells with green fluorescent proteins. *Proc. Natl Acad. Sci. U.S.A.* 95, 6803-6808.
- Longo, D., and Hasty, J. (2006). Dynamics of single-cell gene expression. *Mol. Syst. Biol.* 2, 64.
- Lou, C., Stanton, B., Chen, Y.J., Munsky, B., Voigt, C.A. (2012). Ribozyme-based insulators parts buffer synthetic circuits from genetic context. *Nat. Biotech.* 30, 1137-1142.
- Luker, K. E., Smith, M. C. P., Luker, G. D., Gammon, S. T., Piwnica-Worms, H., and Piwnica-Worms, D. (2004). Kinetics of regulated protein-protein interactions revealed with firefly luciferase complementation imaging in cells and living animals. *Proc. Natl. Acad. Sci. U.S.A.* 101, 12288-12293.
- Mahdavi, A., Segall-Shapiro, T. H., Kou, S., Jindal, G. A., Hoff, K. G., Liu, S., Chitsaz, M., Ismagilov, R. F., Silberg, J. J., and Tirrell, D. A. (2013). A genetically encoded AND gate for cell-targeted metabolic labeling of proteins. *J. Am. Chem. Soc.* 135, 2979-2982.
- Mallam, A.L., and Jackson, S.E. (2011). Knot formation in newly translated proteins is spontaneous and accelerated by chaperonins. *Nat. Chem. Biol.*, 8, 147-153.
- Mansfield, M.L. (1994). Are there knots in proteins? *Nat. Struct. Biol.* 1, 213-214.
- Marcheschi, R. J., Gronenberg, L. S., and Liao, J. C. (2013). Protein engineering for metabolic engineering: current and next-generation tools. *Biotechnol. J.* 8, 545-555.
- Marvin, J.S., Schreiter, E.R., Echevarria, I.M., and Looger, L.L. (2011). A genetically encoded, high-signal-to-noise maltose sensor. *Proteins* 79, 3025-3036.
- Matz, M.V., Fradkov, A.F., Labas, Y.A., Savitsky, A.P., Zaraisky, A.G., Markelov, M.L., and Lukyanov, S.A. (1999). Fluorescent proteins from nonbioluminescent Anthozoa species. *Nat. Biotechnol.* 10, 969-973.
- Mehta, M.M., Liu, S., and Silberg, J.J. (2012). A transposase strategy for creating libraries of circularly permuted proteins. *Nucleic Acids Res.* 40, e71.
- Michnick, S. W., Ear, P. H., Manderson, E. N., Remy, I., and Stefan, E. (2007). Universal strategies in research and drug discovery based on protein-fragment complementation assays. *Nat. Rev. Drug Discovery* 6, 569-582.

- Miyawaki, A., Llopis, J., Heim, R., McCaffery, J.M., Adams, J.A., Ikura, M., and Tsien, R.Y. (1997). Fluorescent indicators for Ca²⁺ based on green fluorescent proteins and calmodulin. *Nature* 388, 882-887.
- Mutalik, V.K., Guimaraes, J.C., Cambray, G., Lam, C., Christoffersen, M.J., Mai, Q.A., Tran, A.B., Paull, M., Keasling JD, Arkin AP, Endy D. (2013). Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat. Methods*, 10, 354-60.
- Nagai, T., Ibata, K., Park, E.S., Kubota, M., Mikoshiba, K., and Miyawaki, A. (2002). A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nat. Biotechnol.* 20, 87-90.
- Narikawa, R., Fukushima, Y., Ishizuka, T., Itoh, S., and Ikeuchi, M. (2008). A novel photoactive GAF domain of cyanobacteriochrome AnPixJ that shows reversible green/red photoconversion. *J. Mol. Biol.* 380, 844-855.
- Nguyen, P. Q. and Silberg, J. J. (2010). A selection that reports on protein–protein interactions within a thermophilic bacterium. *Protein Eng. Des. Sel.* 23, 529-536.
- Nguyen, P. Q., Liu, S., Thompson, J. C., and Silberg, J. J. (2008). Thermostability promotes the cooperative function of split adenylate kinases. *Protein Eng. Des. Sel.* 21, 303-310.
- Nobles, C.L., Clark, J.R., Green, S.I., and Maresso, A.W. (2015). A dual component heme biosensor that integrates heme transport and synthesis in bacteria. *J. Microbiol. Methods* 118, 7-17.
- Nobrega, R.P., Arora, K., Kathuria, S.V., Graceffa, R., Barrea, R.A., Guo, L., Chakravarthy, S., Bilsel, O., Irving, T.C., Brooks, C.L.3rd., and Matthews, C.R. (2014). Modulation of frustration in folding by sequence permutation. *Proc. Natl. Acad. Sci. U.S.A.* 111, 10562-10567.
- Ostermeier, M. (2009). Designing switchable enzymes. *Curr. Opin. Struct. Biol.* 19, 442-448.
- Ostermeier, M., Nixon, A. E., Shim, J. H., and Benkovic, S. J. (1999). Combinatorial protein engineering by incremental truncation. *Proc. Natl. Acad. Sci. U.S.A.* 96, 3562-3567.
- Palmer, E., and Freeman, T. (2004). Investigation into the use of C- and N-terminal GFP fusion proteins for subcellular localization studies using reverse transfection microarrays. *Comp. Funct. Genomics* 5, 342-353.

- Pandey, N., Nobles, C.L., Zechiedrich, L., Maresso, A.W., Silberg, J.J. (2015). Combining random gene fission and rational gene fusion to discover near-infrared fluorescent protein fragments that report on protein-protein interactions. *ACS Synth. Biol.* 4, 615-624.
- Park, S.-Y., Beel, B. D., Simon, M. I., Bilwes, A. M., and Crane, B. R. (2004). In different organisms, the mode of interaction between two signaling proteins is not necessarily conserved. *Proc. Natl. Acad. Sci. U.S.A.* 101, 11646-11651.
- Paschon, D. E. and Ostermeier, M. (2004). Construction of protein fragment complementation libraries using incremental truncation. *Methods Enzymol.* 388, 103-116.
- Paschon, D. E., Patel, Z. S., and Ostermeier, M. (2005). Enhanced catalytic efficiency of aminoglycoside phosphotransferase (3')-IIa achieved through protein fragmentation and reassembly. *J. Mol. Biol.* 353, 26-37.
- Piatkevich, K.D., Subach, F.V., and Verkusha, V.V. (2013). Engineering of bacterial phytochromes for near-infrared imaging, sensing, and light-control in mammals. *Chem. Soc. Rev.* 42, 3441-3452.
- Plaxco, K.W., Simons, K.T., and Baker, D. (1998). Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* 277, 985-994.
- Prasher, D.C, Eckenrode V.K., Ward, W.W., Prendergast, F.G., and Cormier, M.J. (1992). Primary structure of the *Aequorea victoria* green-fluorescent protein. *Gene* 111, 229-233.
- Qian, Z., and Lutz, S. (2005). Improving the catalytic activity of *Candida antarctica* lipase B by circular permutation. *J. Am. Chem. Soc.* 127, 13466-13467.
- Remy, I., Campbell-Valois, F.X., Michnick, S.W. (2007). Detection of protein-protein interactions using a simple survival protein-fragment complementation assay based on the enzyme dihydrofolate reductase. *Nat. Protoc.* 2, 2120-2125.
- Remy, I. and Michnick, S. W. (1999). Clonal selection and *in vivo* quantitation of protein interactions with protein-fragment complementation assays. *Proc. Natl. Acad. Sci. U.S.A.* 96, 5394-5399.
- Resch-Genger, U., Grabolle, M., Cavaliere-Jaricot, S., Nitschke, R., and Nann, T. (2008). Quantum dots versus organic dyes as fluorescent labels. *Nat. Methods* 5, 763-775.
- Rossi, F., Charlton, C. A., and Blau, H. M. (1997). Monitoring protein-protein interactions in intact eukaryotic cells by beta-galactosidase complementation. *Proc. Natl. Acad. Sci. U.S.A.* 94, 8405-8410.

- Rumyantsev, K.A., Shcherbakova, D.M., Zakharova, N.I., Emelyanov, A.V., Turoverov, K.K., and Verkhusha, V.V. (2015). Minimal domain of bacterial phytochrome required for chromophore binding and fluorescence. *Sci. Rep.* 5, 18348.
- Russ, W. P. and Engelman, D. M. (1999). TOXCAT: a measure of transmembrane helix association in a biological membrane. *Proc. Natl. Acad. Sci. U.S.A.* 96, 863-868.
- Sakaguchi, R., Endoh, T., Yamamoto, S., Tainaka, K., Sugimoto, K., Fujieda, N., Kiyonaka, S., Mori, Y., and Morii, T. (2009). A single circularly permuted GFP sensor for inositol-1,3,4,5-tetrakisphosphate based on a split PH domain. *Bioorganic Med. Chem.* 17, 7381-7386.
- Salis, H. M., Mirsky, E. A., and Voigt, C. A. (2009). Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.* 27, 946-950.
- Samuel, P.P., Smith, L.P., Phillips, G.N., and Olson, J.S. (2015). Apoglobin stability is the major governing both cell-free and in vivo expression of holomyoglobin. *J. Biol. Chem.* 39, 23479-23495.
- Segall-Shapiro, T. H., Meyer, A. J., Ellington, A. D., Sontag, E. D., and Voigt, C. A. (2014). A “resource allocator” for transcription based on a highly fragmented T7 RNA polymerase. *Mol. Syst. Biol.* 10, 742-742.
- Segall-Shapiro, T. H., Nguyen, P. Q., Santos, Dos, E. D., Subedi, S., Judd, J., Suh, J., and Silberg, J. J. (2011). Mesophilic and hyperthermophilic adenylate kinases differ in their tolerance to random fragmentation. *J. Mol. Biol.* 406, 135-148.
- Shcherbakova, D. M., and Verkhusha, V. V. (2013). Near-infrared fluorescent proteins for multicolor *in vivo* imaging. *Nat. Methods* 10, 751-754.
- Sheng, J., Ji, X.F., Wang, F., and Sun, M. (2014). Engineering of *Yarrowia lipolytica* lipase Lip8p by circular permutation to alter substrate and temperature characteristics. *J. Ind. Microbiol. Biotechnol.* 41, 757-62.
- Shiba, S., and Schimmel, P. (1992). Functional assembly of a randomly cleaved protein. *Proc. Natl. Acad. Sci. U.S.A.* 89, 1880-1884.
- Shimomura O., Johnson F.H., and Saiga, Y. (1962). Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, *Aequorea*. *J. Cell Comp. Physiol.* 59, 223-239.
- Shis, D. L. and Bennett, M. R. (2013). Library of synthetic transcriptional AND gates built with split T7 RNA polymerase mutants. *Proc. Natl. Acad. Sci. U.S.A.* 110, 5028-5033.

- Shu, X., Royant, A., Lin, M.Z., Aguilera, T.A., Lev-Ram, V., Steinbach, P.A., and Tsein, R.Y. (2009). Mammalian expression of infrared fluorescent proteins engineered from a bacterial phytochrome. *Science* 324, 804-807.
- Sigala, P.A., Crowley, J.R., Hsieh, S., Henderson, J.P., and Goldberg, D.E. (2012). Direct tests of enzymatic heme degradation by the malaria parasite *Plasmodium falciparum*. *J. Biol. Chem.* 287, 37793-37807.
- Stenler, S., Wiklander, O. P., Badal-Tejedor, M., Turunen, J., Nordin, J. Z., Hallengård, D., Wahren, B., Andaloussi, S. E., Rutland, M. W., Smith, C. E., Lundin, K. E., and Blomberg, P. (2014). Micro-minicircle gene therapy: implications of size on fermentation, complexation, shearing resistance, and expression. *Mol. Ther. Nucleic Acids* 2, e140.
- Sulkowska, J.I., Noel, J.K., Ramirez-Sarmiento, C.A., Rawdon, E.J., Millett, J.N., and Onuchic, J.N. (2013). Knotting pathways in proteins. *Biochem. Soc. Transactions* 41, 523-527.
- Sulkowska, J.I., Sulkowski, P., Szymczak, P., and Cieplak, M. (2008). Stabilizing effect of knots on proteins. *Proc. Natl. Acad. Sci. U.S.A.* 105, 19714-19719.
- Sulkowskaa, J.I., Noelb, J.K., and Onuchic, J.N. (2012). Energy landscape of knotted protein folding. *Proc. Natl. Acad. Sci. U.S.A.* 109, 17783-17788.
- Sun, J., Kelemen, G.H., Fernandez-Abalos, J.M., and Bibb, M.J. (1999). Green fluorescent protein as a reporter for spatial and temporal gene expression in *Streptomyces coelicolor* A3(2). *Microbiology* 145, 2221-2227.
- Tainaka, K., Sakaguchi, R., Hayashi, H., Nakano, S., Liew, F.F., and Morii, T. (2010). Design Strategies of Fluorescent Biosensors Based on Biological Macromolecular Receptors. *Sensors* 10, 1355-1376.
- Tchekanda, E., Sivanesan, D., and Michnick, S. W. (2014). An infrared reporter to detect spatiotemporal dynamics of protein–protein interactions. *Nat. Methods* 11, 641-644.
- To, T., Piggott, B.J., Makhijani, K., Yu, D., Jan, Y.N., Shu, X. (2015). Rationally designed fluorogenic protease reporter visualizes spatiotemporal dynamics of apoptosis *in vivo*. *Proc. Natl. Acad. Sci. U.S.A.* 112, 3338-3343.
- Topell, S., Hennecke, J., and Glockshuber, R. (1999). Circularly permuted variants of the green fluorescent protein. *FEBS Lett.* 457, 283–289.
- Viranu, P., Mirny, L.A., Kardar, M. (2006). Intricate Knots in Proteins: Functions and Evolution. *PLoS Comput. Biol.* 2, e122, 1074-1079.

- Wagner, J.R., Zhang, J., Stetten, D.V., Gunther, M., Murgida, D.H., Mroginski, M.A., Walker, J.M., Forest, K.T., Hildebrandt, P., and Vierstra, R.D. (2008). Mutational analysis of *Deinococcus radiodurnas* bacteriophytochrome reveals key amino acids necessary for the photochromicity and proton exchange cycle of phytochromes. *J. Biol. Chem.* 283, 12212-12226.
- Wagner, J.R., Brunzelle, J.S., Forest, K.T., and Vierstra, R.D. (2005). A light-sensing knot revealed by the structure of the chromophore-binding domain of phytochrome. *Nature* 438, 325-331.
- Wagner, J. R., Zhang, J., Brunzelle, J. S., Vierstra, R. D., and Forest, K. T. (2007). High resolution structure of *Deinococcus* bacteriophytochrome yields new insights into phytochrome architecture and evolution. *J. Biol. Chem.* 282, 12298-12309.
- Weissleder, R., and Ntziachristos, V. (2003). Shedding light onto live molecular targets. *Nat. Med.* 9, 123-128.
- Whitehead, T.A., Bergeron, L.M., and Clark D.S. (2009). Tying up the loose ends: circular permutation decreases the proteolytic susceptibility of recombinant proteins. *Protein Eng. Des. Sel.* 22, 607-613.
- Williams, D. J., Puhl, H. L., and Ikeda, S. R. (2009). Rapid modification of proteins using a rapamycin-inducible tobacco etch virus protease system. *PLoS One* 4, e7474.
- Wilson, C. G. M., Magliery, T. J., and Regan, L. (2004). Detecting protein–protein interactions with GFP-fragment reassembly. *Nat. Methods* 1, 255-262.
- Wycuff, D. R., and Matthews, K. S. (2000). Generation of an AraC-araBAD promoter-regulated T7 expression system. *Anal. Biochem.* 277, 67-73.
- Yeates, T. O., Norcross, T.S., and King, N.P. (2007). Knotted and topologically complex proteins as models for studying folding and stability. *Curr. Opin. Chem. Biol.* 11, 595-603.
- Yu, D., Gustafson, W.C., Han, C., Lafaye, C., Noirclerc-Savoye, M., Ge, W.P., Thayer, D.A., Huang, H., Kornberg, T.B., Royant, A., Jan, L.Y., Jan, Y.N., Weiss, W.A., and Shu, X. (2014). An improved monomeric infrared fluorescent protein for neuronal and tumour brain imaging. *Nat. Commun.* 5, 3626.
- Yu, Y., and Lutz, S. (2010). Improved triglyceride transesterification by circular permuted *Candida antarctica* lipase B. *Biotechnol. Bioeng.* 105, 44-50.
- Yu, Y., and Lutz, S. (2011). Circular permutation: a different way to engineer enzyme structure and function. *Trends Biotechnol.* 29, 18-25.

- Yuan, L., Kurek, I., English, J., and Keenan, R. (2005). Laboratory-directed protein evolution. *Microbiol. Mol. Biol. Rev.* 69, 373-382.
- Yuet, K. P., and Tirrell, D. A. (2013). Chemical tools for temporally and spatially resolved mass spectrometry-based proteomics. *Ann. Biomed. Eng.* 42, 299-311.
- Zhao, N., Fogg, J. M., Zechiedrich, L., and Zu, Y. (2011). Transfection of shRNA-encoding minivector DNA of a few hundred base pairs to regulate gene expression in lymphoma cells. *Gene Ther.* 18, 220-224.
- Zhu, B., Wu, G., Robinson, H., Wilganowski, N., Hall, M.A., Ghosh, S.C., Pinkston, K.L., Azhdarinia, A., Harvey, B.R., Sevick-Muraca, E.M. (2013). Tumor margin detection using quantitative NIRF molecular imaging targeting EpCAM validated by far red gene reporter iRFP. *Mol. Imaging Biol.* 15, 560-568.